HEAVY METALS IN RIVER SEDIMENTS

PAUL ANTHONY CHURCH

A THESIS SUBMITTED TO THE COUNCIL FOR NATIONAL ACADEMIC AWARDS FOR THE PARTIAL FULFILMENT OF THE DEGREE OF M. PHIL

APRIL 1986

Department of Science
The Polytechnic of Wales

in collaboration with

The School of Chemistry
The University of Bristol

AKNOWLEDGMENTS

I would like to thank my supervisors for their help and encouragement during the course of this study. I am particularly grateful to Dr K W Burton for his support, guidance, and invaluable assistance also Dr. G Nickless for his encouragement and helpful comment.

I should also like to thank my family for their patience and particularly my wife for her continuous support and encouragement and for typing the manuscript.

HEAVY METALS IN RIVER SEDIMENTS

P.A.CHURCH

ABSTRACT

Samples of sediment material were collected from the River Cynon. The concentrations of heavy metals in the samples were determined by Atomic Absorption Spectroscopy after digestion of the dried sample. A rapid hydrochloric acid/ hydrogen peroxide digestion was chosen after a brief comparison with the more traditional nitric/ perchloric acid digestion.

The results were used in a survey of the heavy metal concentrations over the length of the river and in comparison with surveys of other systems. This showed that while the Cynon contains higher levels of heavy metals than some other similar systems, it is not grossly polluted and contains considerably lower levels of some metals than some mineralised Cornish rivers.

Factor analysis of the data showed that one major factor was controlling the heavy metal distribution at each site. Profiles of these factors were generated using target factor analysis. The evidence suggests that the local geology is the main factor near the source of the river with coal based material determining the profile of heavy metals in the lower reaches.

Pattern recognition techniques were shown to be capable of classifying the four sites and identifying 'outliers' in the data. Similar sites were found to overlap in the principal component analysis but resolution of these sites was improved by utilising principal components that were less important to the overall classification. Magnesium was identified as the feature with the greatest ability to classify the data. Two similar sites identified in the factor analysis were also shown to be similar by the pattern recognition methods. Other data on the Cynon taken at different times was classified as different from the main data categories thus showing the variation of heavy metal concentration with time.

AIMS OF THE STUDY

The Cynon valley is typical of the ribbon development of the South Wales Valleys with industry and urbanisation following the valley floor virtually along its whole length. In the past it contained several mines most of which are now closed. However, there is still a large working pit near Hirwaun which discharges minewater into the Cynon at the rate of about a million gallons per day. Minewaters from other closed pits in the valley are also discharged into the Cynon. There is an operational limestone quarry at Penderyn, light industries and a wire and cable manufacturer at Aberdare and a smokeless fuel plant at Aberaman. There are also the many spoil heaps that always accompanied the mines, not only in the valley itself but also along one of its major tributaries. It would be expected therefore, that the river should show some evidence of such activities.

Heavy metals in the environment are often used as an indicator of anthropogenic impact as they tend to become mobilised by mining and are added to the environment by industry and urbanisation. Much of this heavy metal input finds its way to the river system, some of it directly, such as in minewaters and drainage, some of it indirectly, such as from the leaching of spoil heaps. Within the river itself heavy metals tend to concentrate in the sediments. This process can occur by a number of mechanisms such as precipitation or adsorption. Thus, a study of heavy metals in river sediments can provide a means by which their input into the environment can be monitored.

The aim of this study was therefore to perform a survey of the sediments of the Cynon with a view to determining the major sources of heavy metals which find their way into the river. To this end sediment samples were taken from a number of sites. They were dried and divided into different size fractions. Duplicate samples of one of these fractions was extracted using a hydrochloric acid/hydrogen peroxide digestion. The heavy metals were determined mainly by the use of atomic absorption spectrophotometry.

The data was analysed using factor analysis in order to ascertain the number of factors controlling the heavy metal distribution at each site. Target profiles were also generated for the major factors with the aim of identifying those factors if possible.

Pattern recognition methods were also applied to the data. The purpose of this was to determine if it was possible to distinguish between the sites using these techniques. Having classified the sites, a further aim was to extract from the results any information regarding the importance of individual metals to the classification.

CONTENTS

INTRODUCTION

Almost 2000 years ago (100 AD) Frontinus decreed that "no one shall with malice pollute the waters (Rome's water supply) where they issue publicly. Should anyone pollute them, his fine shall be ten thousand sestertii"(1)

It seems then that it has long been recognised that water pollution is undesirable and that our water supplies are worthy of our protection and for that reason it is necessary to study them in some detail.

Pollution may be of several forms but in this study heavy metals were the focus of attention. There is no universal agreement as to what the term 'heavy metal' actually encompasses and several various, even contradictory definitions are to be found in the literature. The most common definition uses specific gravity as the criterion but even here there is disagreement. In one dictionary heavy metals are defined as being those metals with a specific gravity greater than 4.0 (2), while another (3) uses a value of 5.0 whereas the commonly accepted value appears to be 4.5. Burrel suggested that heavy metals were those metals in the periodic table that are bounded by titanium, hafnium, antimony and bismuth plus selenium and tellurium (4). This is equivalent to the latter definition of heavy metals i.e a specific gravity of greater than 4.5. A novel approach(5) suggests that the term 'heavy metal' should be abandoned in favour of a classification of metals based more upon their solution chemistry rather than a purely arbitrary borderline. The classification includes three

categories which are: 1. oxygen seeking, 2 .nitrogen or sulphur seeking and 3. intermediate or borderline. This study is restricted to eleven of the nineteen metals commonly accepted as 'heavy metals'.

Heavy metals occur naturally in the aquatic environment and are subject to variations due to mineralogical differences, seasonal changes in flow (6) and different pH and dissolved oxygen levels over the length of rivers (7). Superimposed upon these natural fluctuations anthropogenic sources contribute their effects to greater or lesser degrees.

Mining activity is an obvious source of increased levels of many metals in rivers. Copper, chromium, lead, iron, zinc and manganese are all subject to elevated levels in sediments downstream of mining activity (7,8).

Significant pollution is caused by drainage of highways (9,10). A source of increased lead levels has been attributed to proximity to highways and urban centres (11) presumably as a result of automotive fuel combustion.

Domestic and industrial effluents add to the burden of pollution and much effort is made to monitor trace metals from municipal sewage treatment, as apart from the immediate impact upon the receiving river, the treated water may be reused downstream.

Much of the trace metal burden in aquatic systems is carried as particulate matter in the water as opposed to dissolved free metal ion (12). Wollast (13) discussed this concept and showed that metal enrichment took place in the

particulate phase.

Much early work concentrated on determining the total concentration of heavy metals in the system being studied but the last two decades has seen the blossoming of interest in the speciation, or chemical forms of heavy metals. This is due to recognition of the fact that the chemical form of an element is relevant to the fate of that element; e.g. the toxicity of heavy metals seems to be related to the concentration of the 'free' hydrated ion and not to the total concentration. (14,15)

SPECIATION OF HEAVY METALS

While most water quality criteria are still based on total metal concentrations it has become clear that the role of metals in the environment is not necessarily linked to their total concentration. Speciation - a knowledge of the nature and properties of the various chemical forms in which an element occurs - is required. (16,17)

As all living organisms evolved in the presence of all the naturally occurring elements it is not surprising that many are in fact essential to life. Metals which have been identified as essential trace elements are chromium, cobalt, copper, iron, manganese, molybdenum, nickel, tin, vanadium and zinc while cadmium, lead and tungsten are strongly suspected of being so.(17) Above a particular usually very low concentration, however, all these essential elements become highly toxic to most organisms. The toxicity of these metals to aquatic organisms for example, is determined, not by their total concentrations but by the chemical form of the metal.

Most toxicity studies have shown that it is the free hydrated ion that is the most toxic form of a metal whereas strong complexes are normally non-toxic. (15,18)

Chemical interactions are also dependent on metal speciation, and thus the transport of metals in aquatic systems and the role of sediments and suspended particulate matter will depend very largely on the physico - chemical forms of metals.

The complexity of the factors which control

speciation, e.g. pH, redox potentials, adsorption, kinetics, equilibria, presence of complexes, solid/liquid and liquid/air interactions are such that a complete description is not possible although important information may be gleaned from their study.(16)

An example of the effect of one of these factors is demonstrated in Fig. 1 which shows the adsorption of copper II and silver I on amorphous iron oxide as a function of pH. The adsorption increases from a very low percentage to almost total adsorption over a very narrow pH range. This is typical of most hydrolysable metal ions.(19)

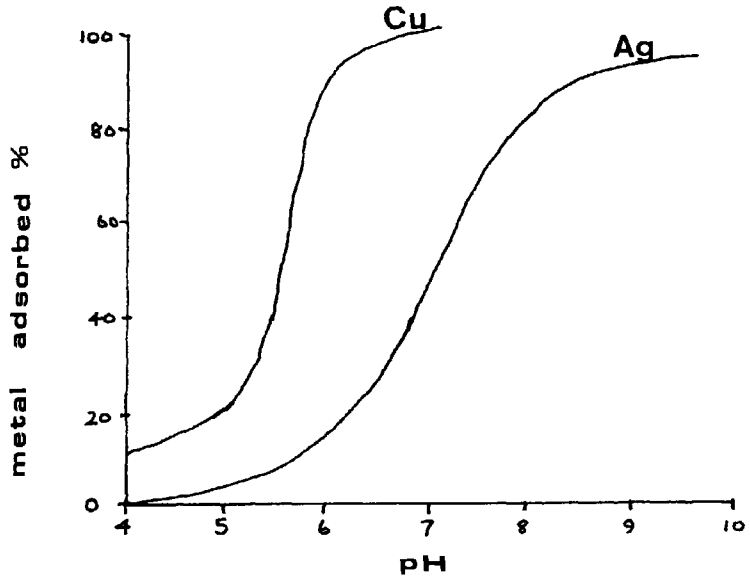


Figure 1. Adsorption of copper(II) and silver(I) on amorphous iron oxide as a function of pH.

The difficulty of ascertaining metal speciation has resulted in a diversity of approaches of analytical methods in speciation studies. In a paper on the subject

Buffle described the requirements and pitfalls of such studies. (16)(Table I)

The very low concentrations of individual species results in enormous problems with contamination and loss of metals by adsorption on vessel walls.(17) Furthermore the various forms are probably not in equilibrium with each other and even if they are, any procedure used may disturb the equilibria and change the speciation. Even the simple act of filtration can alter dissolved oxygen and carbon dioxide concentrations and so affect other equilibria.

A typical classification scheme (20) includes

a) some sort of phase and/or species separation method. The most common practice is to filter through a 0.45 um membrane although other methods such as ultrafiltration can be used. For species separation ion exchange or chelating resins are normally used. The major drawback of these techniques, a criticism applicable to all speciation studies in fact, is that equilibria will be disturbed by removal of a phase or species and alter the concentration being measured.

b) measurement of labile and non-labile species in each fraction obtained from the separation process. The concentrations of metal species are likely to be so low in natural waters that very few analytical techniques are sensitive enough for speciation work.

An ideal technique for measuring a particular species would be the use of ion selective electrodes, as these cause little disturbance of the solution equilibria during

TABLE I

PRINCIPLE	TECHNIQUE	MAIN LIMITATIONS
Detection of atomic and electronic structures	Fluorescence	Broad spectra and interpretation sometimes difficult because of large number of unknown ligands
Potentiometry	Redox/membrane electrodes	Insufficient sensitivity
Voltammetries	DPP,NPP,ASV DPASV	Careful interpretation needed as secondary phenomena (e.g. adsorption) may strongly influence results
Methods based on particle size	Dialysis Gel chromatography	Applicable only to complexes with macromolecules Secondary phenomena (e.g. adsorption may be important
Methods based on chemical reactivity	Solubilisation of solid Competition with ion-exchange resin Extraction Adsorption on solids	Risks of modification of the composition and reactivity of the medium under study
Biological methods	Influence of the free metal ion on the growth of organisms	Influence of speciation on bioassimilation not well known Risks of modification of the medium by compounds excreted by organisms

NPP: Normal pulse polarography
DPP: derivative pulse polarography
ASV: anodic stripping voltammetry
DPASV: differential pulse anodic stripping voltammetry

measurement. Unfortunately ion selective electrodes are far too insensitive for all except the most contaminated of waters.

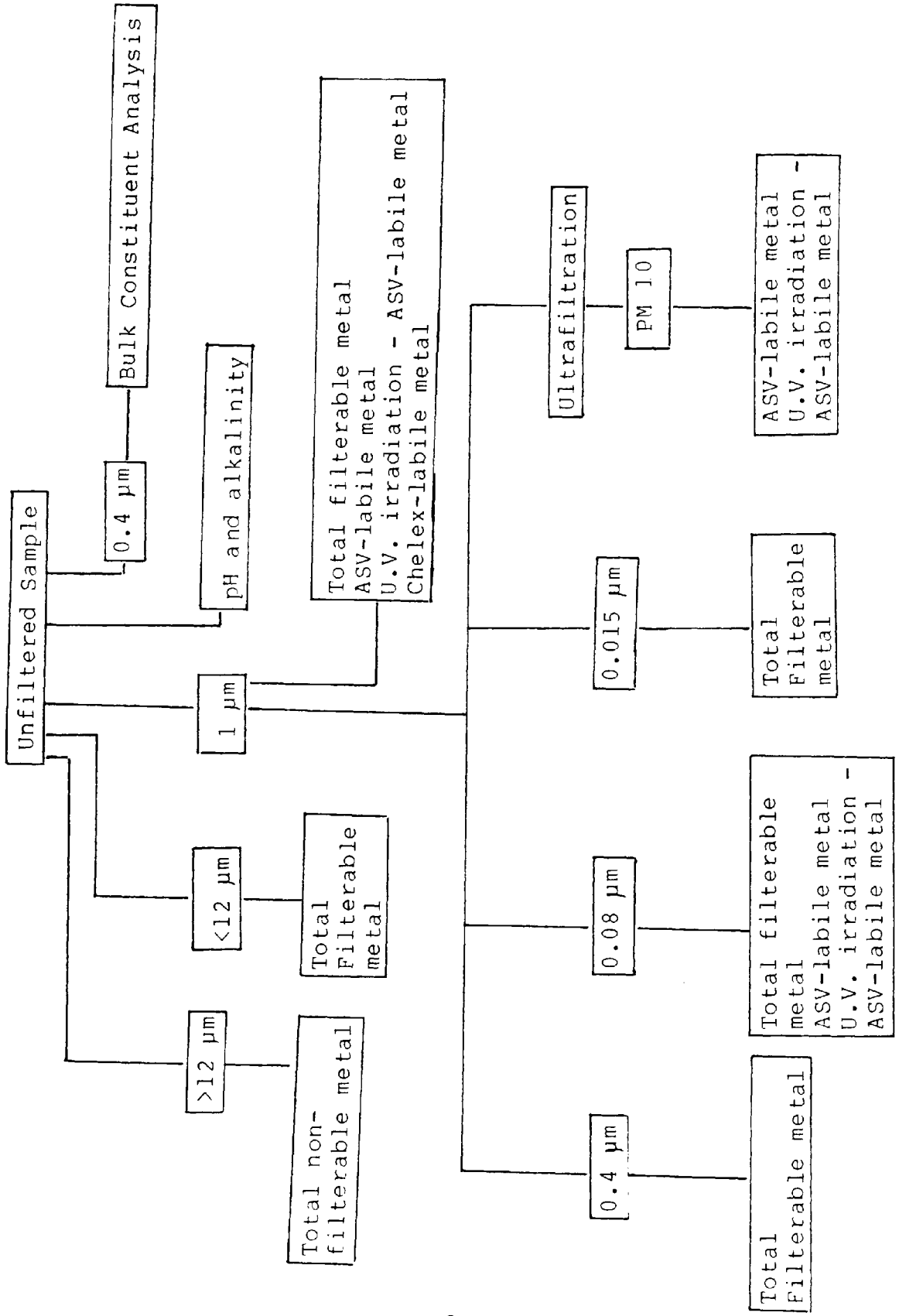
The most common methods used are anodic stripping voltammetry (ASV), ion exchange chromatography, dialysis and bioassay.

A typical speciation scheme is shown in Fig. 2. This scheme was used by Laxen and Harrison (21) in the investigation of sewage effluent and its impact on the receiving river.

They concluded that the major proportion of the heavy metals cadmium, lead, copper and iron were associated with the colloidal and particulate size fractions although manganese did not appear to fit into this pattern. The major part of the cadmium was associated with the smallest size fraction and they suggested that this may be due to cadmium-humic acid complexation.

Mouvet and Bourg (22) studied speciation in the Meuse river and compared their results to values calculated using chemical equilibria constants. They assumed surface sites for adsorption/desorption processes behaved as ligands. While many studies had considered adsorption onto silica, aluminium hydroxide and iron hydroxide as models for adsorption behaviour the authors claimed to be the first to produce quantitative results on **natural** sediments that gave agreement between observed and calculated values. They concluded that adsorption and not precipitation was the controlling mechanism for trace metal concentrations in the river, and that organic and

Figure 2. Speciation scheme - Laxen and Harrison



suspended matter played a major role in copper and lead adsorption.

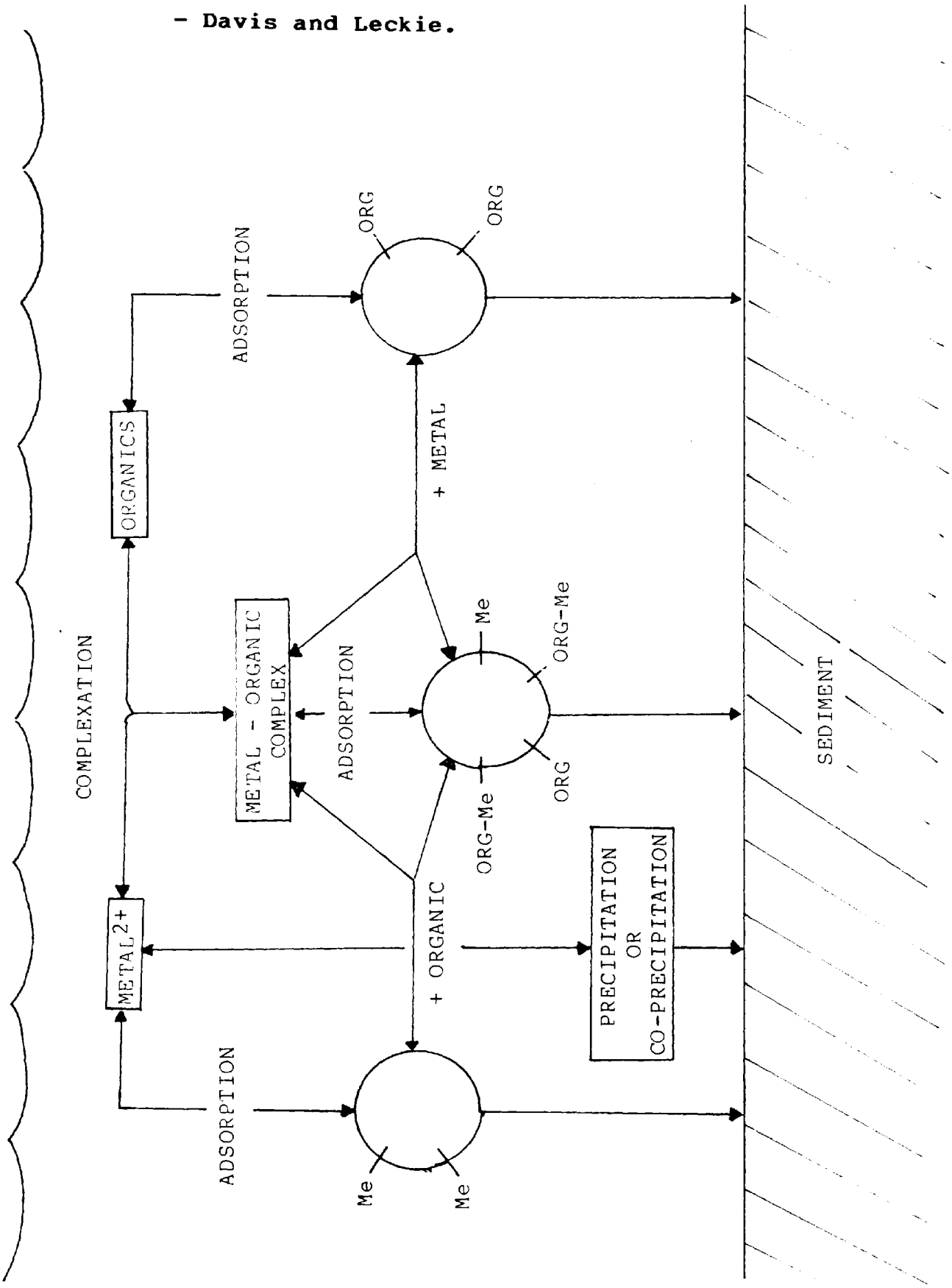
Undoubtedly speciation studies will increase in importance as understanding of trace metals in the environment develops. However without adequate trace metal clean room facilities, techniques for working at the ng dm^{-3} level and the appropriate expertise many trace metal speciation studies must be considered unreliable.(17)

ROLE OF SEDIMENTS

Before entering a discussion of the part sediments play in the transport of trace metals it may be of some use to define what is meant by 'sediments'. A simplistic definition may be 'all material that settles to the bottom of an aqueous medium'. However, when considering heavy metals and other pollutants, adsorption effects play a very important role and so surface area and size of particles must be taken into account. Thus while no unambiguous definition of 'sediments' exists it is generally accepted that for pollution studies, all material that passes a 2mm seive can be usefully used as sediment. Such sediments are often further classified as sand (particle size 2.0 - 0.050 mm), silt (0.050 - 0.002 mm) and clay (<0.002 mm). (United States Department of Agriculture). Chemically the major constituents of sediments are silica and aluminosilicates together with associated iron and manganese hydroxides, trace metals and organics. The relationships between these species is a complex one and will depend on many factors such as pH, redox potential, relative concentrations and the presence of organic and inorganic ligands.

Many studies have shown that trace metals in aquatic systems tend to concentrate in sediments (12,23,24,25). Fig. 3 shows some of the pathways for trace metal deposition in sediments. (19)

Figure 3. Pathways for trace metal deposition in sediments - Davis and Leckie.



There is strong evidence that iron and manganese hydroxides are the major agents for adsorption of heavy metals. (26) However, the presence of organic or inorganic ligands can modify the adsorption processes. Davis and Leckie (19) studied the adsorption of copper (II) and silver (I) ions on amorphous iron oxide. They found that complexing ligands that are not themselves adsorbed e.g. chloride, decrease metal adsorption on oxide surfaces, possibly due to competition between the chloride ion and oxide surface for the metal ion. Complexing ligands that are adsorbed on oxide surfaces may either increase, decrease or have little effect on metal ion adsorption, the particular modification being dependent upon the type of ligand and the comparative strengths of oxide-metal binding and ligand metal binding.

There have been several studies which investigated the effects of changes in the various parameters which control metal adsorption.(27,28,29) Slavek and Pickering (30) investigated the effect of pH on the retention of some heavy metals by clay-fulvic acid mixtures. The results suggested that metal ions would be removed from solution at pH < 6 and tend to increase migration of metal ions into the aqueous phase at pH > 6.

Rendell et al. (31) investigated the adsorption of copper, lead and cadmium onto river sediments and found that significant adsorption occurred even in the presence of extractant solutions such as dilute hydrochloric acid, ammonium acetate solution and sodium citrate solution.

Vuceta and Morgan (32) studied adsorption on silica and came to the following conclusions: copper is extensively adsorbed in the absence of organic complexing ligands but is almost totally released by the presence of such ligands; lead is largely adsorbed at the oxide surface and only organic ligands with a strong affinity for lead disturb this position. These workers also agreed that iron and manganese in aqueous systems are mostly precipitated as oxides both of which adsorb significant amounts of trace metals.

Benjamin and Leckie (33) suggested a model for metal-ligand surface interactions during adsorption. They viewed such a system as consisting of two phases.

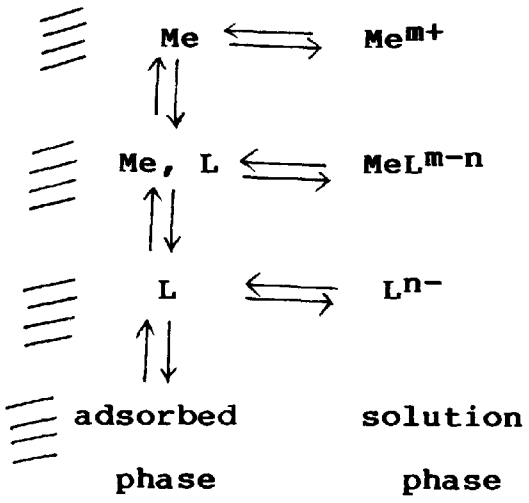


Figure 4. Metal-ligand interactions - Benjamin and Leckie.

They classified the adsorbed complexes as either metallike or ligandlike and developed this concept, the main assumptions of which are shown in Table II.

In a later paper (34) the same authors looked at the adsorption behaviour of cadmium on oxide surfaces and investigated the effects of various ligands on that behaviour. They also used the model above to make inferences about the stereochemistry of adsorbed complexes. They concluded that complexed metal ions can adsorb to oxide surfaces with either the metal or ligand bonding directly to the surface. Adsorption appeared to be a function of speciation in solution, adsorbent concentration and pH but for a given complex they tentatively suggested that the adsorption behaviour is independent of the adsorbent. If the metal end of the complex is adsorbed, then the percentage complex adsorbed as a function of pH is similar to that of the free metal but shifted to a higher pH. If the ligand end of the complex is adsorbed, then as ligand concentration increases metal adsorption increases in one pH range but decreases in a higher pH range.

Thus sediments play an extensive role in the removal of metal ions from the aqueous phase. The evidence seems to show that a large proportion of the heavy metal load in the system is carried by particulate and sedimentary material but that changes in pH and the presence of chelating agents can cause mobilisation of metals.

TABLE II

assumption	most likely deviation	effects on predicted sorption patterns
1) all MeL_x complexes adsorb identically	MeL_1, MeL_2 .... have different binding strengths to surface	analysis more complicated and quantitative predictions altered some; qualitative pattern probably not altered much.
2) fractional sorption of a species independent of species concentration	surface site non-homogenicity	for ligandlike complexes, pH_c decreases with decreasing free metal concentration; region of nonmonotonic relation between fractional metal sorption and ligand concentration appears; for metallike complexes, adsorption edge shifts to higher pH with increasing metal concentration.
	surface sites limiting	not likely to occur for metallike complexes; decreases maximum sorption at low pH for ligandlike complexes.
	shifting IEP with ligand concentration	negligible effect on metallike complexes; for ligandlike complexes, pH_c decreases with increasing ligand concentration; metal sorption decreases some at low pH, but eventually attains 100% removed; region of nonmonotonic relation between metal sorption and ligand concentration appears.
3) complexes sorb by only one mechanism	complexes can sorb in either metallike or ligandlike fashion	analysis involves combining analyses for two limiting cases; no major effect.

pH = pH at which fractional adsorption of free metal equals that of complexed metal

IEP = isoelectric point

HEAVY METAL ANALYSIS IN WATER AND SEDIMENTS

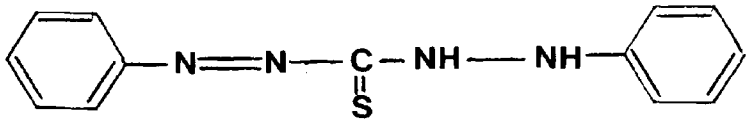
Trace metals in natural waters tend to be present at very low levels unless there is some dramatic input such as mining activity.(5,8) However, the cumulative effect of trace metals can make even very low levels hazardous and thus the desirability of measuring such levels has lead to the development of a number of appropriate analytical techniques.

Preconcentration of the samples is one method of attempting to measure very low levels of metal ions, although the associated difficulty of high reagent blanks may present some problems. One method of preconcentration is simple evaporation but this may lead to loss of volatile metals such as mercury. Another possibility is to use ion exchange resins. The cations present in a large volume of water may be absorbed onto a small amount of an appropriate resin. The metal ions may then be eluted using a small volume of dilute acid.

The well established photometric methods will almost certainly require preconcentration in order to provide metal ion concentrations high enough for the limited sensitivity of these methods.

An example of a commonly used reagent for photometric heavy metal determination is dithizone (28,29) which forms coloured chelate complexes with many metal ions.

Figure 5. Structural formula of dithizone



By controlling such factors as pH and the solvent used for extraction it is possible to use dithizone for selectively determining many trace metals. The dithizone complexes with the **free** metal ions and therefore other species present may not be detected. Alternatively, as some organo-metallic complexes also complex with dithizone they may interfere with the analysis.

Prior treatments include boiling with acid to dissolve suspended particulate matter and refluxing with concentrated H_2SO_4 and $KMnO_4$ to oxidise organo-metallic complexes.

The water sample is treated with appropriate reagents to mask interferences from other metals and its pH adjusted. The dithizone is dissolved in $CHCl_3$ or CCl_4 the choice being fixed by the metal being determined. It is then added to the sample and shaken vigorously. The absorbance of the complex formed is measured for the sample and for calibration samples which have undergone the same procedure.

Atomic absortion spectrophotometry (AAS) may be used to determine directly heavy metals in water(37) but as with the previous method the sensitivity is not normally high enough to detect metals at the levels usually found

and so preconcentration may be necessary. The advantage of this method is that complications due to interference from other metals do not normally arise or if they do are usually simply dealt with.

The principle of atomic absorption is similar to that of molecular absorption, that is the species under investigation absorbs light with an energy associated with a particular electronic transition. In molecular spectroscopy a whole series of transitions are possible associated with vibrational and rotational excited states in the molecules giving rise to spectra with close lying bands. In solutions however, broad band spectra are produced as small energy changes are absorbed by interactions with solvent molecules.

In atomic spectra none of these complexities are present. The spectrum of one element will consist of a number of narrow absorption lines each line corresponding to one specific electronic transition.

These narrow absorption bands, commonly about 0.005 nm wide, present a slight problem. Most absorption spectroscopy utilises a continuum light source with a monochromator to isolate the region of the spectrum under investigation. The limit of such a system is a band of the order of 0.05 nm wide. Thus only a very small proportion of this band is absorbed in atomic absorption resulting in a low absorbance $[-\log_{10} I_t/I_o]$ and poor sensitivity.

The problem was solved in the early 1950s by Sir Alan Walsh (38) who realised that if a line source rather than a continuum light source was used not only would higher

sensitivity be achieved but also interference from other absorbing lines would be much reduced as overlap of very narrow lines becomes less likely. The line source most commonly used is the hollow cathode lamp.

The cathode is cylindrical and made from the analyte element or an appropriate alloy. The lamp is filled at low pressure (6×10^2 Pa) usually with neon. The discharge occurs between the cathode and a small anode and it is centralised by mica strips inside the lamp. The end window is usually of silica as glass absorbs U.V. light.

The sample is atomised by introducing an aerosol into the flame using a nebuliser which aspirates the sample. The flame used is normally of air-acetylene: however elements (e.g. aluminium) which form refractory oxides need the hotter nitrous oxide-acetylene flame. A long thin flame is used as this improves sensitivity. Light from the lamp is focussed through the flame and then further focussed into the monochromator which isolates light of the appropriate wavelength. On exiting from the monochromator it falls onto a photomultiplier and is converted to an electrical signal.

The power supply to the lamp is modulated giving an A.C. signal while light emitted from the flame will produce a D.C. signal, this allows discrimination between light from lamp and light from the flame.

Interferences

Because each element absorbs at only a few wavelengths and over a very narrow range, overlap - two elements absorbing at the same wavelength is not normally a problem for most analyses. Molecular absorption may cause some problems but can usually be overcome by operating a continuum source alternated with a line source. The two signals can be distinguished by appropriate electrical design and so isolating the absorbance due to the analyte atoms.

Certain chemical interferences are possible due to the presence of thermally stable compounds in the flame, e.g. silicate can suppress calcium, magnesium and iron absorbance but the addition of a releasing agent such as lanthanum to both standards and samples greatly improves absorbance by forming a more stable compound with silicate than the other elements.

In hot flames some elements may ionise and so reduce the absorbance of the atoms. This is not necessarily a problem if the same processes occur in the standards, but in a complex matrix other easily ionisable elements may shift the equilibria in the sample, increasing its absorbance while not doing so in the standard. In this study this particular effect may occur with aluminium in the hot nitrous oxide - acetylene flame. It is overcome by adding a large excess of an easily ionisable element such as potassium to both the sample and standard producing a large excess of electrons in the flame and so suppressing ionisation of the analyte atom:

$$M(g) \rightleftharpoons M^{+}(g) + e$$

Flame Emission Spectroscopy - (FES)

When solutions containing heavy metals are aspirated into a flame, some of the atoms produced are thermally excited. A proportion of these atoms lose energy associated with the excited state by the emission of light. The intensity of light is dependent upon (albeit indirectly) the concentration of the metal present in the solution.

The techniques and requirements of FES are similar in many respects to AAS. Both may be performed on the same instrument but there are some differences. In AAS, transitions that originate from the ground state are the most significant whereas in emission many more lines are possible. Thus in order to separate an analyte line from lines of other elements a high resolution monochromator is necessary.

A simplified instrument known as a flame photometer may be used for a few elements - in particular some of the alkali and alkali earth metals. These few elements emit at wavelengths longer than most others and so can be atomised and excited in low temperature flames such as air/natural gas. As so few elements are involved a simple filter may be used instead of a monochromator and a photocell can take the place of a photomultiplier. The main elements that can be determined in this way are

lithium, sodium, potassium, calcium and barium. In this study calcium was determined using flame photometry.

Anodic Stripping Voltammetry - (ASV)

This is a very sensitive technique for determining trace metals in water and because of its sensitivity widely used for natural waters particularly in speciation studies.(39,40)

The basis of ASV is to be found in polarography, the principle of which utilises the reduction of metal ions at a dropping mercury electrode. The current which flows is determined by the rate of diffusion of the dissolved metal ions which in turn is proportional to the metal ion concentration. Furthermore, the potential at which the metal ions are reduced is characteristic of the ion itself and not of its concentration making selective determinations possible.

The dropping mercury electrode is useful as its continuous renewal means that uncontaminated mercury is used throughout the analysis. Also hydrogen ions have a high overvoltage at mercury surfaces preventing their reduction and interference. Unfortunately the fluctuation of the current measurements as the mercury drop grows and eventually falls limits the sensitivity of this technique.

ASV overcomes this problem by using a stationary electrode e.g. a hanging mercury drop. The electrode is maintained at a negative potential while the ion of interest is reduced to the metal which forms an amalgam

with the mercury. The amount of ion deposited in this way is still proportional to the concentration in the solution. The rate of deposition can be increased by stirring. With stirring stopped the potential is slowly increased. When the applied potential reaches the half-wave potential for the metal being determined the dissolved metal is oxidised back to its ion which diffuses back into the solution. This is the "stripping" stage,and results in a rapid change of current. The height of the current peak is proportional to the original ion concentration.

A problem of this method is that quite broad current peaks can occur due to the slowness of the diffusion of metal ions out of the drop of mercury. However, other electrodes e.g. graphite impregnated with wax or glassy carbon electrodes overcome this problem.

A further refinement of this method is to superimpose small pulses of potential on the linearly increasing potential during the stripping process. The current is measured before and at the end of each pulse. Electronic manipulation of the two signals can eliminate background noise and increase sensitivity even further. This technique is known as differential pulse stripping voltammetry - DPSV.

These techniques only determine the dissolved ionic species and not metal ions that are complexed. This is turned to advantage however in speciation studies where the determined metal ion concentration is termed 'ASV labile'.

ANALYSIS OF DATA - a literature survey

Analysis of large data sets such as were produced in this study, a 140 x 11 data matrix, is obviously an extremely difficult task. The process can be simplified by taking the means of elemental concentrations at each site but data reduction of this type is almost bound to result in loss of information.

In the last decade or so a branch of chemistry called chemometrics has been established which is capable of dealing with such large data sets. Chemometrics may be defined as the utilisation of mathematical and statistical methods for handling and interpreting chemical data. Several techniques, but two in particular, factor analysis and cluster analysis have been found to be useful in this respect. (41,42,43)

The use of factor analysis and cluster analysis in the elucidation of environmental data has been established by several workers (44,45,46,47) but these methods have not yet found widespread application in the environmental field.

Elemental analyses of environmental samples are however particularly amenable to factor analysis (42) for the following reason. The total amount of element i in sample k (x_{ik}) can be expressed as

$$x_{ik} = \sum_{j=1}^{n} c_{ij}\, q_{jk}$$

where c_{ij} is the relevant concentration of element i in source j, q_{jk} is the amount of source j in sample k and the sum is taken over n significant sources of that element - i.e. each source behaves as a factor. The use of factor analysis and cluster analysis together provides a powerful tool for extracting information from large data sets.

One of the earliest uses of factor analysis was an attempt by Blifford and Meeker (48) to interpret the elemental composition data obtained on atmospheric samples by the National Air Sampling Network during 1957 - 61. They used a principal components analysis together with a Varimax and non-orthogonal rotation. Their attempt met with little success however possibly due to the fact that only 13 elements were determined and the sampling covered 30 cities and hence there were too wide a variety of sources for meaningful interpretations to be made.

Dunn et al (45) used the pattern recognition methods SIMCA and KNN to classify polychlorinated biphenyls in environmental samples. The commercial names of these chemicals in the United States are Aroclor 1242, 1248, 1254 and 1260. They each consist of a very complex mixture of around 125 isomers each, Aroclor being characterised by a chromatographic profile. While visual inspection of chromatographs may enable classification of the Aroclors, problems may be incurred in classifying an unknown source, or sources that may have been modified from use.

Dunn et al were able to classify the Aroclors and a

sample of transformer oil of uncertain origin using the pattern recognition techniques stated above. They further suggested that it would be possible to use the results obtained to determine isomer specific variation that had occurred in a sample through modification by mixing with other Aroclors, differential partitioning in water or sediments, weathering, dilution, biomodification or any combination of these.

Jellum et al (49) were able to distinguish between normal brain tissue and tumour brain tissue using capillary gas chromatographic profiles.

Dunn et al (50) identified the - CHCl group as opposed to a - CCl_2 group as being significantly involved in the mutagenicity of a halogenated hydrocarbon using a principal component model of pattern recognition.

Hopke et al (51) applied factor analysis to identify the sources of atmospheric aerosols in the Boston area and were able to identify 6 sources, soil, sea salt, oil-fired power plants, motor vehicles, refuse incineration and an unknown magnesium /selenium source. Hopke applied target transformation factor analysis to atmospheric samples collected in St. Louis USA (52). A component related to potassium was identified and found to have a significant impact on three samples. These were identified as having been collected at a site and time when a firework display had been in progress. Hopke emphasised the point that the analysis was able to identify a perturbation that affected only three samples. These were removed from the data set and the remaining data was reanalysed. Five

factors were identified, sulphate, motor vehicles, paint pigment, refuse incinerators and the fifth factor as either coal flyash or soil, the similarity in their elemental profiles making distinction between them very difficult.

A number of other workers have also successfully applied factor analysis for the identification of sources of atmospheric pollution. (46,53,54)

Little work has been done on the application of multivariate analysis on aquatic systems. Hopke (47) applied such techniques to chemical and physical data of lake sediments. Concentrations of 15 elements were determined. Particle size distribution, organic matter, water depth above sample and grainsize distribution data obtained by other workers were included in the analysis. Hierarchical cluster analysis and factor analysis were used to interpret the data leading to the conclusion that five factors were required to describe the variation in the nature of the sediments and the processes acting on them. These were coarse-grained source material, glacial till source material, surface area per unit volume, active sedimentation and wave/current action. Furthermore, the analysis not only provided useful information about the lake sediments but indicated areas that required additional information before a more complete picture of the system could be developed.

Bobee et al (55) also used factor analysis and a hierarchical clustering on eight parameters describing water quality of the Yamaska river. They were able to

compare the variability of these parameters. They examined the clustering of the sampling stations and were able to reduce the number of stations required to monitor water quality by a factor of four without appreciable loss of information.

Tada et al (56) used a cluster analysis according to Euclidian distance to classify Japanese rivers by heavy metals in their sediments. They measured the concentrations of copper, zinc, lead and cadmium extracted from the sediments of 94 Japanese rivers taken from both upper and lower reaches of each river. The distributions of the metals were compared with respect to the upper and lower reaches. A cluster analysis was performed to determine the degree of similarity between rivers based on heavy metals in the sediments of the lower reaches of the rivers. This enabled the authors to classify Japanese rivers into three groups. The first group had a high content of the four metals measured. The second group had metal concentrations lower or only slightly higher than the average of the upper reaches of all the rivers. The third group had very high metal concentrations.

A pollution index suggested by Nishida et al (57) was determined for the lower and upper reaches of the rivers and a value of this index was chosen as the critical value for pollution of the rivers. A comparison of this index for the lower reaches with the cluster analysis showed a high degree of correlation. This enabled the authors to classify the rivers as follows Group A - polluted with heavy metals, pollution index in the range 3 - 7, mainly

rivers which flow through regional cities; Group B - non polluted, pollution index 0.1 - 2.5; Group C - heavily polluted, index in the range 4 - 27, containing rivers that flow through large cities or receive mine drainage.

Hoban (58) used factor analysis on data obtained about PAH (polyaromatic hydrocarbons) levels in sediments of the Severn Estuary. He concluded that only one factor controlled PAH levels in the estuary. He was able to identify this factor by TFA (target factor analysis). Using an iterative procedure in which a row (site) vector consisting of one value of unity and the remainder of zeros he predicted that the organic content of the sediment was the factor controlling the PAH levels. Application of TFA using the organic content of the sediments as the target confirmed this hypothesis. He also applied several pattern recognition techniques to the data and was able to obtain information regarding minor perturbations in the data which the factor analysis was unable to resolve.

ANALYTICAL PROCEDURE

Heavy metal concentrations were obtained on Cynon river sediments using the following procedure:

1. Sample collection.
2. Drying and fractionation.
3. Digestion.
4. Measurement.
5. Determination of organic content of sediment.

The Cynon Valley

The Cynon River was chosen as the sampling area for several reasons:

It is a river typical of the urban and industrial development of the South Wales Valleys and would therefore be expected to show some signs of pollution from a number of sources e.g mining activity. It is conveniently close geographically and it was known that good sampling sites were available and accessible. Figs 6,7

The river flows through an essentially urban, light industrial area for most of its course. The Cynon Valley in common with the other South Wales valleys is a typical example of a ribbon development with urbanisation and industry following the valley floor along its length. The river rises in the millstone grit of the southern edge of the Brecon Beacons, passes through a limestone band near

Figure 6. Geology of sampling area.

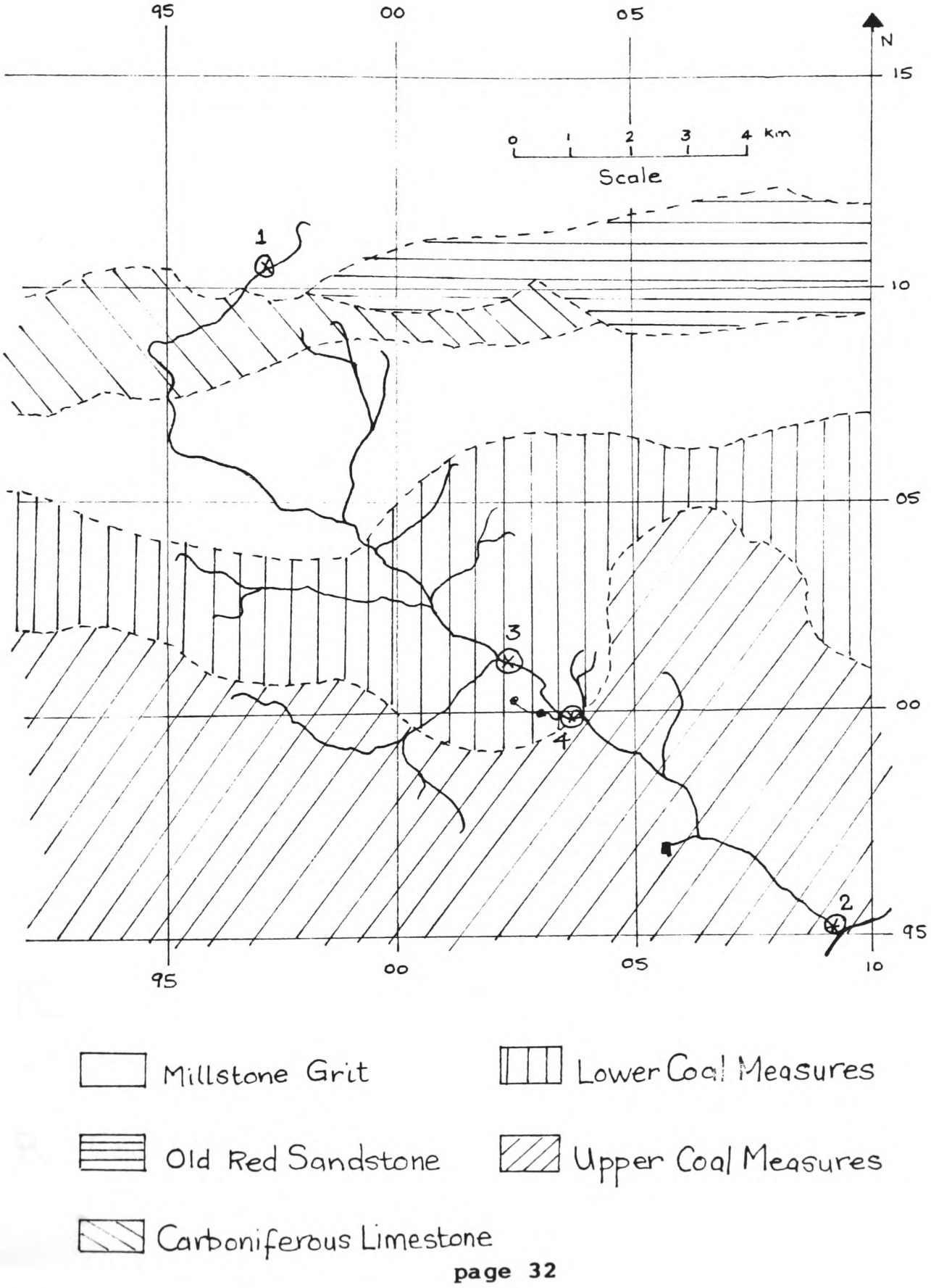
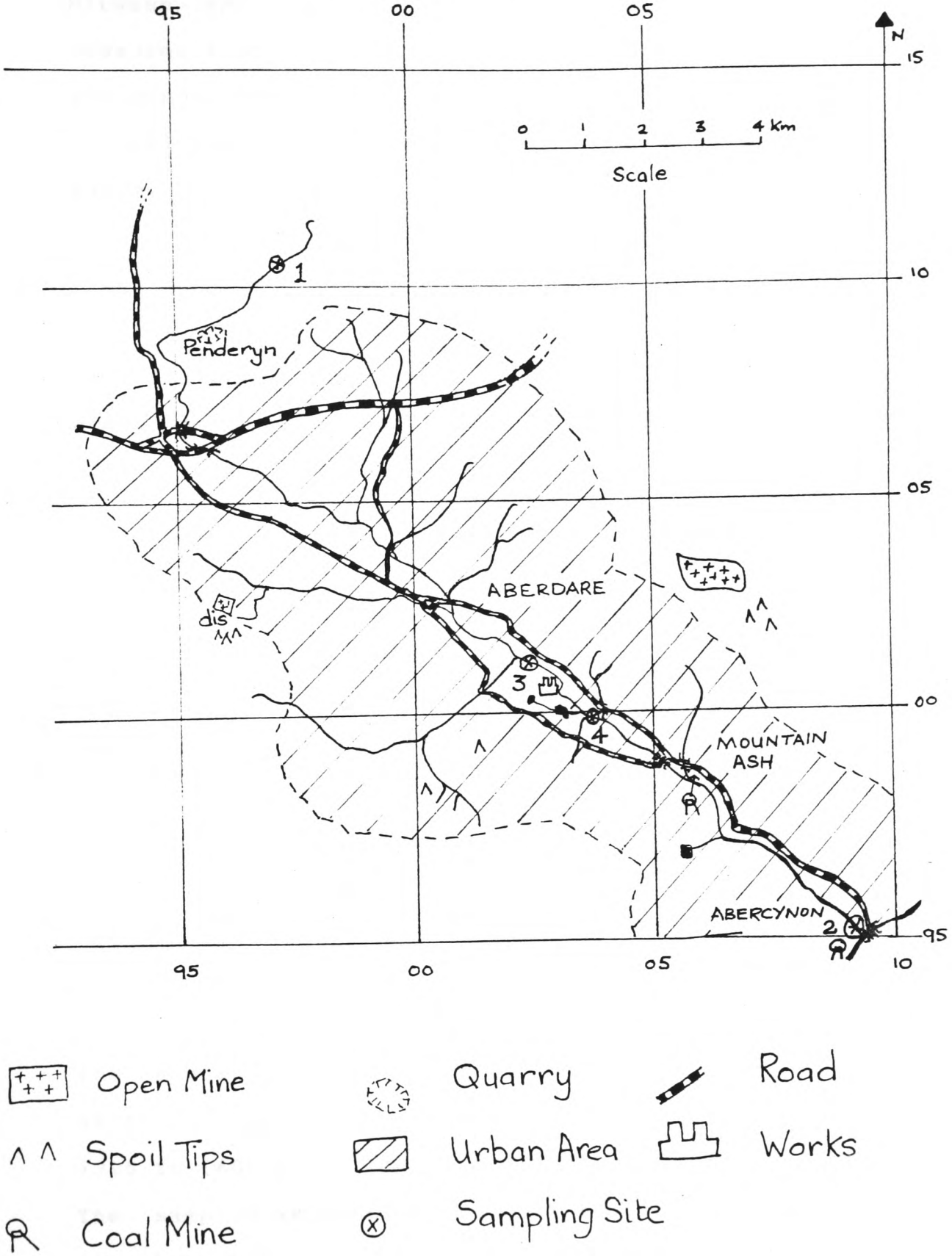


Figure 7. Geography of sampling area.



Penderyn and a further band of millstone grit above Hirwaun and then through the lower and middle coal measures from Hirwaun to its confluence with the Taff at Abercynon.(59,60)

All the major sewage and industrial effluent of the valley is treated at the Cwm Farm water treatment works Pontypridd for discharge into the River Taff so the Cynon receives little of this type of effluent. However it is subject to all the other urban impacts such as storm water runoff, spoil tip seepage, old mineworkings and the atmospheric input associated with an urban/industrial area.

Sample Collection

Samples were collected from four sites on the River Cynon. A polythene scoop was used to obtain the sediment which was then transferred to a labelled polythene bag for transport back to the laboratory. Twenty five samples were collected at each of sites 1 and 2 whilst twelve samples were collected at both sites 3 and 4.

Site 1.

This was on open moorland about 0.5 kilometers from the source of the Cynon (grid ref SN 973106). The stream at this point is known as Nant Cadlan, it rises at about 1325 ft (400 m) on the southern slopes of Cader Fawr 480m. The samples were collected over about 70 m. of the stream on the southern side of the minor road which crosses the

stream. The rock in this area is essentially millstone grit.

Site 2.

This site was at a sports centre near the confluence of the river Cynon and the river Taff, at grid reference ST 082955. The river is bounded by playing fields on the western bank and overgrown natural vegetation on the opposite bank. Two busy roads on either side of the river run parallel to and a few hundred metres from the river.

Site 3.

The junction of the Nant Aman Fach and the Cynon about two kilometres downstream of Aberdare at grid reference SO 020012 was the location for site 3. At this stage the Cynon has passed through a limestone area near Penderyn at which there is a quarry, the urban and light industrial area of Hirwaun and the town of Aberdare. Along its course it will have passed the sites of several old mineworking and spoil tips. The Nant Aman Fach rises on the ridge dividing the Rhondda Fach and Cynon valleys, flowing through the village of Cwmaman (old mineworkings and spoil tips) before joining the Cynon at the site of a new light industrial estate.

Site 4.

Below a footbridge approximately one kilometre above the village of Mountain Ash, grid reference ST 037997, site four was adjacent to some playing fields which are bounded by two busy main roads, one on either side of the

river. About one kilometre further upstream the river flows through the Phurnacite works, a large plant operated by the National Coal Board for the production of smokeless fuel from coal. The main effluent from this plant is treated by a bio-oxidation process, after settling in a lagoon, before discharge into the Cynon. Also the plant uses large quantities of coal dust and involves partial distillation of the coal with resultant atmospheric problems so it is likely that it may have a significant impact on the river.

Drying and Fractionation Procedure

All the glassware used at this and subsequent stages was cleaned using the following protocol:

1. wash in detergent
2. several rinses in tap water
3. soak in Decon 90
4. rinse in deionised water
5. soak in 10% nitric acid
6. rinse in deionised water

On arrival at the laboratory each sample was transferred from its polythene bag into a beaker. All the samples were dried in an oven at 105°C for 24 hours.

The dried samples were separated into fractions by dry sieving through a stack of stainless steel sieves placed on a sieve shaker for 10 minutes each. The sieve mesh sizes were 2 mm; 0.5 mm; 0.25 mm; 0.106 mm and 0.075

mm. Only the material that passed the 0.075 mm sieve was used for the heavy metal determinations as this should contain the major burden of heavy metals. (10,61,62)

Digestion

For total elemental analysis the sediments would normally be dissolved in a suitable acid mixture such as hydrofluoric/perchloric acid . The hydrofluoric acid is necessary to dissolve the silicates present but the apparatus used must therefore be of an inert material such as platinum or Teflon. For heavy metal analysis an acid mixture such as nitric/perchloric can be used but requires quite a long digestion period, typically 24 hours. A more rapid method which uses a hydrochloric acid hydrogen peroxide mixture (63) was considered in comparison with a nitric acid perchloric acid digestion.

Two factors were investigated :- 1)the correlation with the nitric acid perchloric acid method and 2) the minimum mass of sample needed to obtain a satisfactory determination.

Four samples of the < 0.075 mm material were digested using both the hydrochloric acid hydrogen peroxide and nitric acid perchloric acid methods. For the digestion using hydrochloric acid and hydrogen peroxide masses of each sample from 0.2 g to 1.0 g were placed in boiling tubes together with 5 cm^3 of analar concentrated hydrochloric acid and 1 cm^3 of analar 30 % w/v hydrogen

peroxide. After the initial reaction had subsided the tubes in a metal rack were placed on a hot plate for 2 hours and maintained at temperature close to 100°C.

For the alternative digestion method 3 cm^3 of analar concentrated nitric acid and 2 cm^3 of analar concentrated perchloric acid were added to 1 g of each sample in boiling tubes. These were also heated on a hot plate in a metal rack at about 90°C for 24 hours.

After the appropriate digestion time all the samples were filtered through Whatman No 44 filter paper into a 25 cm^3 graduated flask. The tubes were rinsed with deionised water and then the filter papers and contents were also rinsed with deionised water.

Zinc, copper, nickel,chromium, lead and cadmium were determined for these samples. The results are shown in Appendix A.

It was concluded from these results that the more rapid hydrochloric acid/hydrogen peroxide method compared favourably with the much slower nitric acid/perchloric acid method and so this method of digestion was adopted for all further work.

Using the hydrochloric acid/hydrogen peroxide method, in the range 0.21 - 1.0 g, there appeared to be no optimal sample mass. Therefore a mass of 0.25 g was selected for all the digestions as it was economical on sample and simplified the calculation of metal concentration per gramme of solid.

0.25 g of each sample was digested using the method

described above. Each digestion was performed in duplicate. Each 25 cm^3 graduated flask receiving the filtrate after digestion contained 1 cm^3 of analar lanthanum chloride (670 g dm^{-3} $LaCl_3.7H_2O$) solution, 1 cm^3 of analar potassium chloride (96 g dm^{-3} KCl) and 1 cm^3 analar nitric acid. The lanthanum chloride was added as a releasing agent for calcium and magnesium determinations. The potassium chloride was added to eliminate ionisation interferences in the hot nitrous oxide - acetylene flame in the determination of aluminium. The acid was added to prevent hydrolysis of the samples during storage.

Blank determinations were carried out at the same time as sample determinations following exactly the same procedure as for the samples.

The solutions were transferred to dry polythene bottles which had undergone the same washing treatment as the glassware.

Measurement

Calcium was determined using a Corning flame photometer. All other metals were determined using a Varian atomic absorption spectrophotometer with an air - acetylene flame except aluminium, for which a nitrous oxide - acetylene flame was used. The instrument parameters for the metal determinations are listed in Appendix B.

In some cases the concentration of metal fell outside

the working range of the AAS instrument at the wavelength being used. In such cases the samples were diluted by a factor of ten together with the blanks. The lanthanum, potassium and acid concentrations were also adjusted accordingly.

The results obtained from the AAS gave metal concentrations in $\mu g\ cm^{-3}$ of solution. After adjustment for the reagent blanks the concentration of the metal per gramme of dry sediment was obtained using the following formula:

$$\text{concentration of metal M } (\mu g\ g^{-1}) = \frac{\text{conc. in digested soln.} \times 25}{\text{mass of sediment used.}}$$

The results are listed in Appendix C.

Organic Content

The organic content was estimated by loss on ignition. Samples were placed into porcelain crucibles which had been washed, dried and heated in a muffle furnace at 550°C for 2 hours. Crucible and samples were dried at 100°C for one hour, then weighed before being placed in the furnace at 550°C for one hour. The difference in mass on cooling (loss on ignition) was used as an estimate of organic matter content (64). The results are listed in Appendix D.

ANALYSIS OF DATA - Theory

Data obtained from aquatic and other environmental systems is notoriously difficult to interpret due to its complexity. It often consists of the measurement of a large number of variables for many individual samples thus producing a large data matrix which traditional techniques handle with difficulty. In the last decade or so, a methodology has evolved, by which meaningful analyses of the data can be made. This is a discipline known as chemometrics.

Chemometrics may be defined as the utilisation of mathematical and statistical methods for handling and interpreting chemical data. The methods used are essentially those of multivariate analysis, which have in fact long been used in other studies involving large data sets such as the social sciences. It is only comparatively recently that their usefulness in interpreting chemical data has been recognised.

The techniques available fall broadly into two groups:
1)PATTERN RECOGNITION 2)FACTOR ANALYSIS

1) PATTERN RECOGNITION

Pattern recognition techniques are usually subdivided into two groups a) unsupervised analysis and b) supervised analysis.

a) unsupervised analysis.

With large data sets, any inherent structure or

regularity is most unlikely to be recognised using normal statistical methods. Unsupervised analysis attempts to discover if any systematic structure exists within the data. Searching for regularity within the samples (which have been characterised by measurement of a number of variables) is usually known as CLUSTER ANALYSIS.

b) supervised analysis

These techniques may be applied in situations where the data set is known to be composed of specified groups or categories. The analysis attempts to produce a classification rule that will identify members of known categories. The data used to generate the classification of these groups is called the TRAINING SET. Data used to test the validity of the classification is called the EVALUATION SET and data for which the classification is unknown but is to be determined using the training set is called the TEST SET.

2) FACTOR ANALYSIS

In this study the data matrix obtained contains several rows. These are the measurements of the concentrations of a number of heavy metals for each sample collected from one site. These columns of the data matrix contain the measurements of the concentration of a particular heavy metal for all the samples.

$$[D] = \begin{array}{l} \\ \text{rows} = \text{samples} \end{array} \begin{array}{c} \text{columns = heavy metals} \\ \begin{bmatrix} d_{11} & d_{12} & d_{13} & \cdots & d_{1c} \\ d_{21} & d_{22} & d_{23} & \cdots & d_{2c} \\ d_{31} & d_{32} & d_{33} & \cdots & d_{3c} \\ \vdots & \vdots & \vdots & & \vdots \\ d_{r1} & d_{r2} & d_{r3} & \cdots & d_{rc} \end{bmatrix} \end{array}$$

Each data point is represented by the symbol d_{ik} this being the data point associated with the ith sample and the k th metal. Factor analysis views the data point d_{ik} as a linear sum of the metal concentrations from several sources (factors) each factor being weighted differently. The data point d_{ik} is assumed to have the form

$$d_{ik} = m_{i1}l_{1k} + m_{i2}l_{2k} + \cdots m_{in}l_{nk}$$

$$= \sum_{j=1}^{n} m_{ij}l_{jk}.$$

where m_{ij} is the metal concentration in source j and l_{jk} is the relative loading of source j. The number of terms, n, in the above expression is the number of factors or sources that contribute to the metal concentration in each sample.

The goal of factor analysis is to identify the <u>number</u> of factors n, and if possible to identify each factor - in this study the source of the heavy metals.

The data matrix [D] can be thought of as the product of two matrices [R] and [C] where

$$[R]_{abstract} = \begin{matrix} \\ s \\ a \\ m \\ p \\ l \\ e \end{matrix} \begin{matrix} \text{factor} \\ \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ r_{r1} & r_{r2} & \cdots & r_{rn} \end{bmatrix} \end{matrix}$$

$$[C]_{abstract} = \begin{matrix} \\ f \\ a \\ c \\ t \\ o \\ r \end{matrix} \begin{matrix} \text{concentration} \\ \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1c} \\ c_{21} & c_{22} & \cdots & c_{2c} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ c_{n1} & c_{n2} & \cdots & c_{nc} \end{bmatrix} \end{matrix}$$

$$[D] = [R]\,[C]$$

Decomposition of the matrix in this way is a purely mathematical device. The row matrix [R] and the column matrix [C] have no physical meaning and are therefore called abstract matrices.

While the matrices themselves have no physical meaning the number of factors have a real significance.

The procedure by which the abstract matrices are extracted involves a technique called eigenanalysis. This produces a set of eigenvectors and an associated set of eigenvalues. Each eigenvector represents a factor

while the associated eigenvalues are a measure of the importance of each factor. If there were no experimental error the number of eigenvectors would correspond exactly to the number of controlling factors, but because of the inevitability of experimental error the resultant eigenanalysis always produces the same number of eigenvectors as there are rows in the data matrix. The eigenvectors are ranked in order of ability to account for variation in the data and are calculated to give the solution to the equation

$$[D] = [R]\,[C]$$

where

$$[C]_{abstract} = \begin{array}{c} \\ \text{f} \\ \text{a} \\ \text{c} \\ \text{t} \\ \text{o} \\ \text{r} \end{array} \begin{array}{c} \text{concentration} \\ \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1c} \\ c_{21} & c_{22} & \cdots & c_{2c} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ c_{n1} & c_{n2} & \cdots & c_{nc} \end{bmatrix} \end{array}$$

The first factor accounts for the highest proportion of the variation in the data matrix and is represented by the first row in the abstract column matrix. The last row represents the least important being associated with the smallest eigenvalue.

These factors can be divided into two sets, one consisting of n factors which account for real measurable differences in data while the second set c - n (where c is

the total number of factors, equal to the number of rows in the abstract column matrix) is associated with the experimental error. If the error in the data is known a number of methods exist to determine the number of factors. If the error is not known, there are other techniques for extracting the factor size. The raw data matrix is the sum of two matrices, the pure data matrix [D^*] (a matrix consisting of data points which are free of experimental error) and an error matrix [E].

$$[D] = [D^*] + [E]$$

Factor analysis of the error matrix can produce a number of error functions which can be utilised in an empirical manner to determine the factor space.

Pind

This is an empirical function that becomes negative when the factor space is overspanned.

Imbedded Error Function

$$IE = \left[n \frac{\sum_{j=n+1}^{c} \lambda_j^o}{rc(c-n)} \right]^{1/2}$$

Normally the IE reaches a minimum at the number of factors responsible for the data, increasing or remaining constant as the increasing number of factors account for more and more of the error in the data matrix.

Indicator Function

$$IND = \frac{RE}{(c-n)^2} \qquad \text{where} \quad RE = \left[\frac{\sum_{j=n+1}^{c} \lambda_j^{o}}{r(c-n)} \right]^{1/2}$$

As for the IE , the IND reaches a minimum when the correct number of factors is used. This is an empirical function and described by Malinowski as more sensitive than IE.(42)

Factor Loading

Successive eigenvectors involve increasing random error and so the mean of the loadings should be close to zero for error eigenvectors.

Transformation

Having determined the number of factors further stages are necessary to obtain an insight into the nature of the factors. Transformation is the next stage and involves converting the abstract factors into real factors by utilising a transformation matrix [T]. The process can be described mathematically as

$$[D] = \left\{[R]_{abs}\ [T]\right\}\ \left\{[T]^{-1}\ [C]_{abs}\right\}$$
$$= [R]_{trans}\ [C]_{trans}$$

To perform this transformation two approaches are possible

a) target transformation - tests potential factors one at a time

b) rotation - transforms abstract matrices into other abstract matrices which may be more amenable to interpretation. This approach has been widely used in the behavioural sciences for many years (65).

Transformation by target testing can identify real factors provided that appropriate data on the suspected sample is available. The testing can use either row or column designees and can be summarised for a row matrix as

$$[R]\ T = R_{predicted}$$

The [R] is the row matrix from the principal component analysis based on n factors. The vector T is obtained from a least squares process involving the target and the principal component analysis solution. A comparison of the test vector with the predicted vector R determines whether the test vector is a real factor or not. If the test vector is reasonably similar to the predicted vector $R_{predicted}$, then R_{test} can be adjudged a real factor, if sufficiently dissimilar the test vector can be rejected as a real factor.

Another type of target testing utilises a target vector consisting of unity and zeros. This target vector undergoes an iterative process to produce a predicted vector which may be compared with real test vectors.

Hopke (52) suggested that this may provide a means to develop elemental profiles for sources were no previous measurements have been made. He developed such profiles

in resolving data taken in the Regional Air Pollution Study, St. Louis Mo. Hoban (58) also used this procedure to generate a profile of the factor controlling PAH in the Severn Estuary.

RESULTS

The means and the range of values for each metal were determined for each site. (Table III). At site 1 the Cd, Cr, Pb, Mn and Zn concentrations were significantly higher than at any of the other three sites. This site is on open moorland, near the source of the river and before it has received drainage from any industrial or domestic source and is therefore free from any obvious pollution. The high means for these metals at this site is therefore rather surprising.

Site 3 is the next site about 12 km downstream from site 1, it has a much higher Ca level than that for site 1 while the values for Cd, Cr, Pb, Mg, Mn, Ni and Zn are all considerably lower. A further 3km downstream was site 4. The concentrations of heavy metals found here were very similar to those found at site 3. Except for Mn which showed a very slight increase between sites 3 and 4, all values were slightly lower.

The last site was a further 7 km downriver, site 2, about 250 m above the confluence of the Cynon and the Taff. The levels of Al, Ca, Mn, and Zn were all lower than at the previous site whilst Cu, Pb and Ni showed slightly higher levels. Mg levels at this site were larger than at any other sample site on the river, the minimum concentration at site 2 (1800 $\mu g\ g^{-1}$) exceeding the maximum value for all the other sites by some 340 $\mu g\ g^{-1}$ (site 1 maximum 1460 $\mu g\ g^{-1}$)

TABLE III

	Al	Ca	Cd	Cr	Cu	Fe	Pb	Mg	Mn	Ni	Zn
SITE 1											
MEAN	7427	8431	28.1	20.4	39.6	37369	186	606	13658	95	2208
MAX	17500	21500	102.0	64.0	150.0	163000	623	1460	48900	429	8000
MIN	1790	2130	0.5	4.0	7.9	2500	23	119	175	8	61
SITE 3											
MEAN	9492	15621	1.0	12.0	79.5	45188	105	33	1031	38	239
MAX	11100	21400	2	15	159	62500	165	36	1420	77	293
MIN	7400	9400	0	6	47	37000	58	28	750	23	177

TABLE III (cont)

	Al	Ca	Cd	Cr	Cu	Fe	Pb	Mg	Mn	Ni	Zn
SITE 4											
MEAN	7658	11971	1.7	6.7	48.9	32792	89	24	1066	23	196
MAX	9600	32800	3	12	78	52000	184	34	3400	35	238
MIN	5600	3900	0	5	36	21500	44	17	400	16	135
SITE 2											
MEAN	5010	6925	2.1	11.4	94.2	46463	114	2271	701	41	159
MAX	7590	9200	3.0	17.0	316.0	74000	268	2900	960	55	215
MIN	3400	4420	1.0	5.0	52.0	26500	79	1800	340	31	128

TABLE IIIA

ONE WAY ANALYSIS OF VARIANCE USING SCHEFFE TEST
(* DENOTES SITES SIGNIFICANTLY DIFFERENT AT THE 0.050 LEVEL)

METAL	SITES 1/2	SITES 1/3	SITES 1/4	SITES 2/3	SITES 2/4	SITES 3/4
Al	*	*		*	*	*
Ca		*	*	*	*	
Cd	*	*	*			
Cr	*	*	*			
Cu	*	*			*	*
Fe					*	
Pb	*	*	*			
Mg	*	*	*	*	*	
Mn	*	*	*			
Ni	*	*	*			
Zn	*	*	*			

TABLE IIIB

POOLED VARIANCE ESTIMATE

METAL	SITES 1/2		SITES 1/3		SITES 1/4		SITES 2/3		SITES 2/4		SITES 3/4	
	T VALUE	2-TAIL PROB.	T VALUE	2-TAIL PROB.	T VALUE	2-TAIL PROB.	T VALUE	2-TAIL PROB.	T VALUE	2-TAIL PROB.	T VALUE	2-TAIL PROB.
Al	4.92	0.000	-2.97	0.004	-0.21	0.834	-20.77	0.000	-10.21	0.000	5.45	0.000
Ca	1.98	0.051	-6.26	0.000	-2.37	0.020	-17.09	0.000	-4.39	0.000	1.85	0.070
Cd	7.73	0.000	5.67	0.000	5.29	0.000	8.33	0.000	2.49	0.015	-2.76	0.008
Cr	5.15	0.000	3.44	0.001	5.54	0.000	-0.86	0.394	7.78	0.000	8.13	0.000
Cu	-6.63	0.000	-4.93	0.000	-0.89	0.375	1.42	0.159	4.79	0.000	5.33	0.000
Fe	-1.99	0.050	-1.28	0.203	0.87	0.386	0.49	0.629	5.13	0.000	5.99	0.000
Pb	3.85	0.000	3.06	0.003	3.86	0.000	0.91	0.364	3.64	0.001	2.46	0.018
Mg	-28.14	0.000	9.34	0.000	9.09	0.000	38.16	0.000	36.87	0.000	7.85	0.000
Mn	7.88	0.000	5.42	0.000	5.16	0.000	-9.08	0.000	-3.23	0.002	-0.35	0.725
Ni	3.82	0.000	2.88	0.005	3.51	0.001	1.95	0.055	12.55	0.000	5.95	0.000
Zn	7.78	0.000	5.27	0.000	5.15	0.000	-12.72	0.000	-6.69	0.000	5.02	0.000

Trends in metal concentrations downstream

Figures 8 - 11 show the variation of the mean concentrations of the heavy metals on moving downstrean. Each figure contains all the heavy metal concentrations that are of the same order of magnitude.

Cd, Cr, Pb, Mn, Ni, Zn.

All of these metals show a similar trend in that they all have a high concentration at site 1.The concentrations decrease downstream although there is a slight rise towards the mouth of the river at site 2 for most of them.

Cu.

This metal shows an overall increase in concentration on moving downstream. There is a slight drop from site 3 to site 4 but nevertheless the trend is for an increase in concentration.

Al, Ca.

These metals behave similarly in that there is an increase from site 1 to site 3 and then a steady decline down stream.

Mg

This metal shows a most unusual trend compared to the others in that its concentration decreases over the first three sites then rises dramatically at the lowest site.

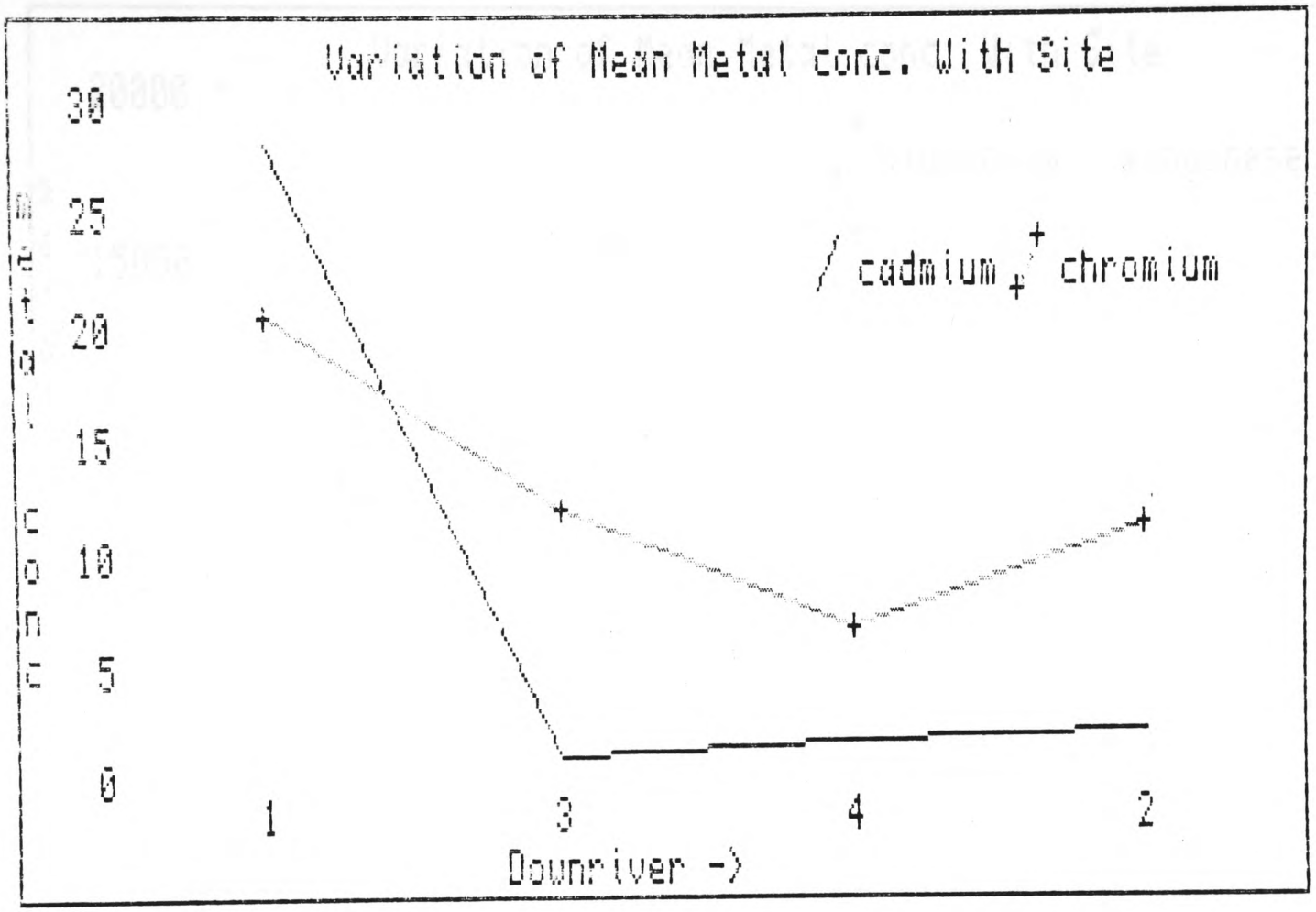

Variation of Mean Metal conc. With Site
30
25
20
15
10
5
0
1
3
4
2
Downriver ->
cadmium
chromium

Figure 8.

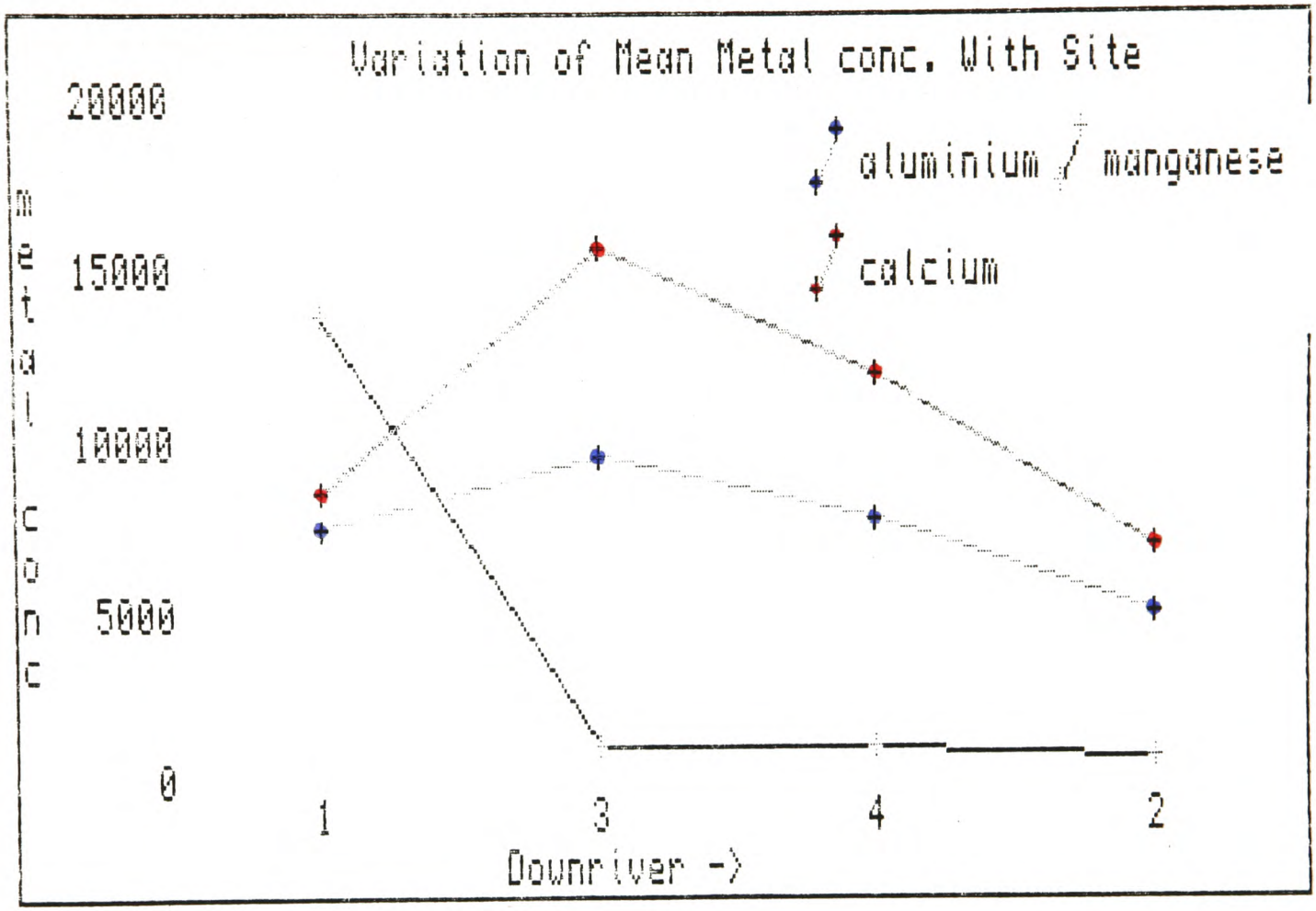

Variation of Mean Metal conc. With Site
20000
15000
10000
5000
0
metal conc
aluminium / manganese
calcium
1
3
4
2
Downriver ->

Figure 9.

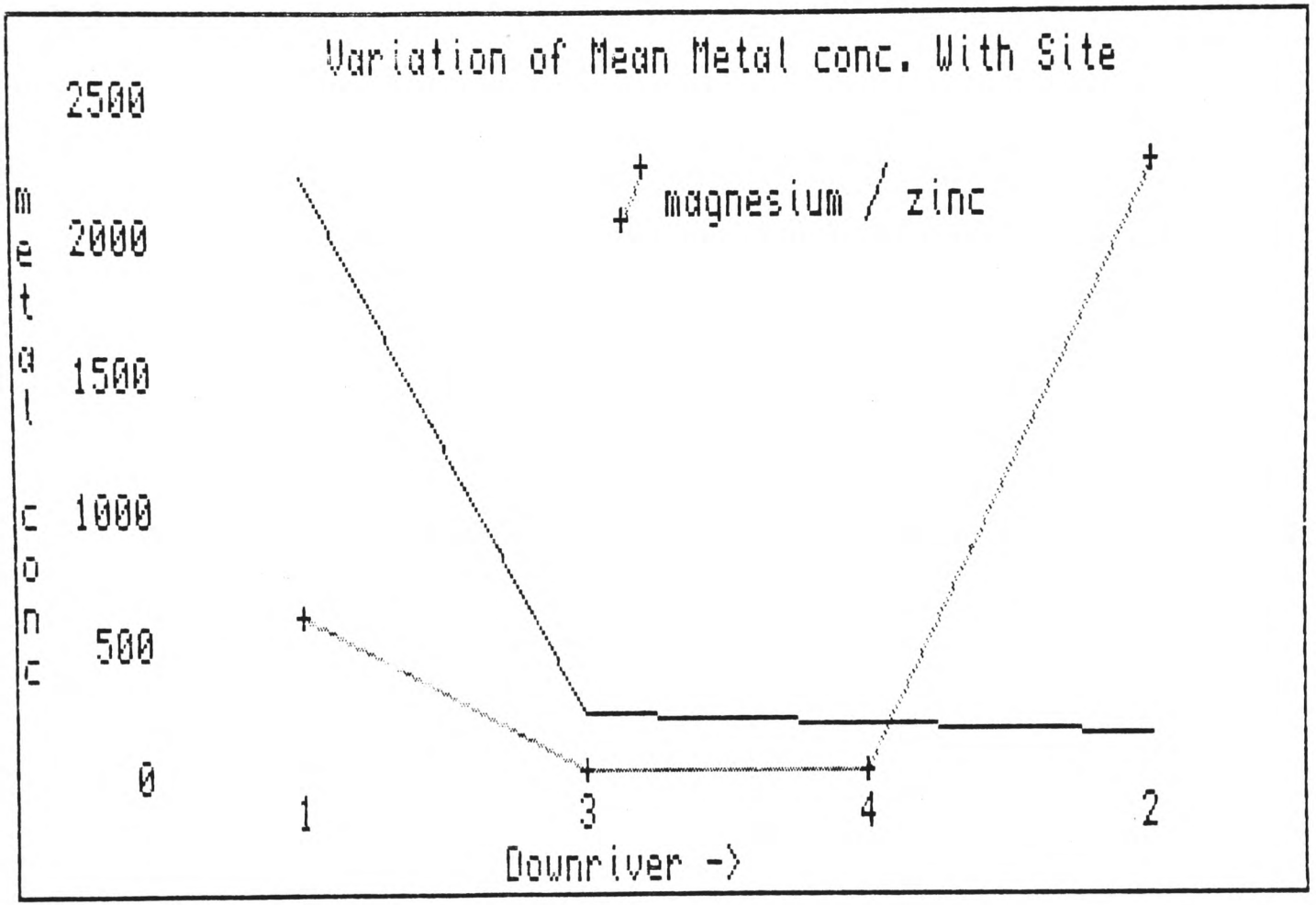

Variation of Mean Metal conc. With Site
2500
2000
1500
1000
500
0
metal conc
magnesium / zinc
1
3
4
2
Downriver ->

Figure 10.

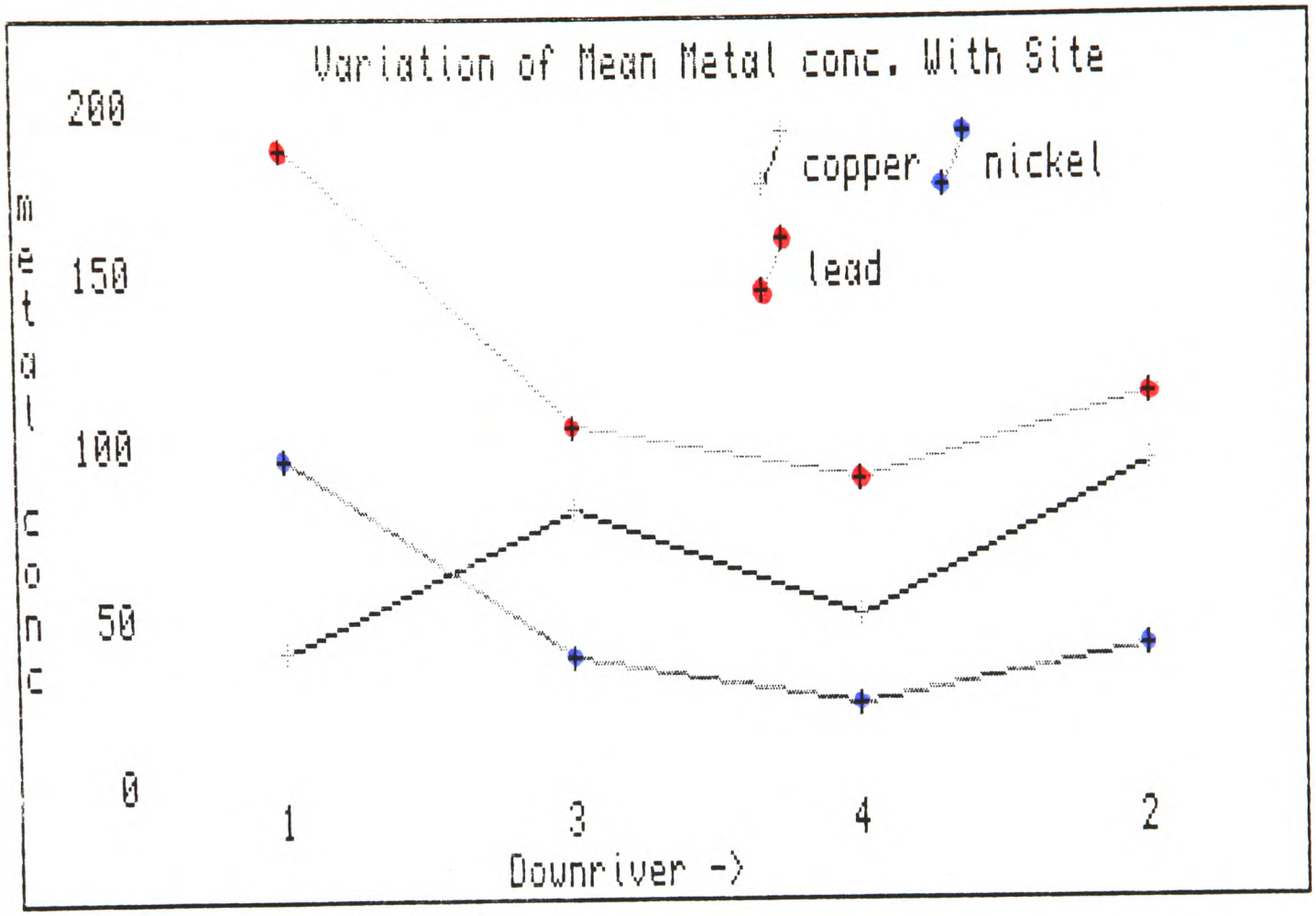

Variation of Mean Metal conc. With Site
200
150
100
50
0
metal conc
copper
nickel
lead
1
3
4
2
Downriver ->

Figure 11.

Fe.

The Fe concentrations were of the same order at all the sites, about 40 000 $\mu g\ g^{-1}$.

It is difficult to make definitive statements about the degree of anthropogenic pollution of a river without a very extensive survey of the background levels of heavy metals in the catchment area and indeed, a more extensive survey of sites within the river, than was possible in this study. Furthermore comparisons with surveys of other river systems can only be of a very generalised form due to such factors as differences in the geology of the systems and in the methods of determining heavy metals in sediments.

However Table IV compares the mean values taken over all samples at all sites for some of the heavy metal concentrations in the Cynon with those determined for other systems.

Tada et al (56) surveyed sediments of 92 Japanese rivers , the concentrations of Cu, Zn, Pb and Cd were determined by extracting the sediment with 0.5M hydrochloric acid for 24 hours. It might be expected therefore that given a particular sample sediment the heavy metal levels may be lower than for the hydrochloric acid / peroxide digestion used in this study. Two values are shown in the table, those for the upper reaches (uncontaminated by waste water etc.) and those for the lower reaches (subject to industrial and urban impact).

Table IV

	Japanese rivers			Cornwall rivers		Antarctica
Metal	River Cynon	Upper reaches	Lower reaches	River Fowey	River Carnon	Shea Sisters Lake
Cd	10.6	0.2	0.62	3	5	
Cu	66	7.1	22.8	50	1650	53
Fe	41000			33000	57000	24800
Pb	133	7.7	14.3	100	470	55
Mn	5200			740	1005	967
Zn	870	17.3	61.5	150	1080	77

The mean values for the Cynon are considerably higher than those quoted by Tada et al. As stated above some of this difference may be attributed to the difference in the method of determination of the metals. However the differences are so large that it is reasonable to conclude that the Cynon contains considerably higher concentrations of Cu, Zn, Pb and Cd than the mean value of the 92 Japanese rivers.

Lyons et al (66) analysed 25% HCl leachable Fe, Mn, Cu, Pb and Zn from the sediments of two lakes in Antarctica. The figures quoted in Table IV are for Shea Sisters Lake, a 'freshwater' lake. Once again the Cynon values are considerably higher although as before some of the difference may be attributed to the methods used. The values for this lake, presumably as unaffected by anthropogenic sources as is possible, are generally higher than those for the Japanese rivers. In the case of Cu the values for the Lake and the Cynon are very similar although for the other methods they are between 2 and 10 times higher.

Aston et al (6,67) investigated the sediments of four Cornish rivers. One of these, the Fowey was chosen by the authors to provide a baseline for their study as it is free of mineralisation due to mining activities and flows through open moorland or agricultural land. The geology and soil type of the Fowey catchment area is not dissimilar to that of the upper reaches of the Cynon. The other three all flow through mining areas where extensive mineralisation has occured. The values for one of these,

the Carnon, is also shown in Table IV as it had the highest values of the three. Aston et al used a nitric / perchloric acid digestion to determine the metal concentrations and it would therefore be expected to produce similar results on similar sediments as the hydrochloric acid / hydrogen peroxide digestion used in this study.

The Cd, Mn and Zn concentrations were all higher in the Cynon than in the Fowey by a factor of between 3 and 7 while Cu, Fe and Pb values were of the same order.

Comparing the Cynon with the Carnon, Cd and Mn were still higher in the Cynon while Fe and Zn were of the same order. However, Cu and Pb were considerably higher in the Carnon.

From this evidence it would seem reasonable to conclude that the heavy metal concentrations in the Cynon are higher than in sediments from both unpolluted and polluted sources. Considering the industrial past of the Cynon, its passage through a mining valley, the presence of spoil heaps in its catchment area, a working limestone quarry in its upper reaches and extensive urbanisation it does not seem as heavily polluted as perhaps might be expected. The values at site 1 are particularly relevant to this conclusion. The site is on open moorland away from the influence of both agricultural and domestic sources of pollution. No gross effect (such as individual items of rubbish) could be found on a visual inspection above the sampling site. The only possible source of imput in addition to the natural geology is a

minor road which crosses the stream. The road is however little used and could not reasonably be expected to have anything but a very minor impact. Despite this site 1 provided the highest mean values for Cd, Cr, Pb, Mn, Ni and Zn. It also provided the highest values for an individual sample for Al and Fe as well as for the metals above.

On the evidence available it would seem that the values at site 1 are due to the inherent geology of the immediate vicinity. The lower values downstream are possibly due to dilution by sediments containing lower concentrations of these metals. The increase in Ca values between sites 1 and 3 suggest an additional input which could be accounted for by the limestone area and quarry at Penderyn. The increasing Cu concentrations downstream may be accounted for by a steady input along the length of the river while the dramatic increase in Mg concentration at the lowest site suggests an individual source.

The mean organic concentration at each site increases steadily downstream as shown in Fig 12.

Figure 12. Mean organic content(%) at each site as determined by ignition loss.

Site 1	Site 2	Site 3	Site 4
9.00	12.83	18.25	31.82

This is probably due to an increase in the amount of coal based material in the river sediment. This postulate is supported by the observation that the sediment collected from the lower two sites were darker in colour than from the upper two. Indeed many of the samples collected at the lower site were black. This coal based material may have been deposited some time ago when colliery washings were discharged into the river. An alternative and more likely possibility is that a small but steady seepage of fine material finds its way to the river from the Phurnacite plant. In fact large quantities of the fine coal dust used at the plant is stored not very far from the river. This could account for the quite dramatic increase in organic content of the sediment of the two sites downstream of the plant.

ANALYSIS OF EXPERIMENTAL DATA

The data was investigated using Factor Analysis and by Pattern recognition Techniques.

Factor Analysis

The analysis was performed using the computer program 'TARGET' by E.R. Malinowski on an Apple IIe microcomputer. The aim of the factor analysis was to determine the number of factors contributing to the heavy metal load in the sediments at each site and if possible to identify a source profile for each factor.

The first stage in this process was to obtain the eigenvalue analysis (EVA) at each site. A uniqueness test was then performed to identify any samples that are 'outliers' and so may have an undue influence on the factor analysis. These can be identified by low values of the 'SPOIL' function. Any offending samples were removed from the data matrix and the eigenvalue analysis was then repeated. The number of factors contributing to the heavy metal concentration at each site was determined.

An attempt was then made to generate the profile of each factor at each site by providing a target vector consisting of unity and zeros. Twelve of these targets were generated, the first being unity and eleven zeros, the second a zero, unity and ten zeros, the third two zeros, unity and nine zeros etc. The target factor analysis provides a predicted vector from each of these

target vectors which was then used as the new target vector. This iterative process was repeated until no significant difference occurred between the target and predicted vectors. These vectors should represent the ratios of heavy metals in the profile of each factor (52).

Site 1.

Table V shows the results of the eigenvalue analysis. Eight rows of the data matrix had already been removed due to incomplete data. The data was column normalised before the analysis because of the different orders of magnitude in the data. The uniqueness test Table VI showed that rows 15 and 16 had low SPOIL values but it was felt that these were not low enough to justify removal from the data matrix.

The PIND fuction indicated 2 factors, the IND fuction showed a minimum at 3 factors while the IE was inconclusive. Table VII shows the cumulative percentages of the eigenvalues of the first five eigenvectors at each site. This shows that 5 factors are needed to account for 99% of the variation in the data at site 1 although only 2 factors are required to account for about 95%.

When using the target factor analysis (TFA) program, if one factor was specified all the initial target vectors iterated to the same predicted target vector. If two factors were specified then the initial twelve target vectors on iteration produced five distinct target vectors. One of these contained negative values which

Table V

Site 1

SUMMARY OF EVA RESULTS USING (COLUMN) CORRELATION

NF	EV	RE	IE	IND	PIND
1	10.7269838	3511.23366	1013.60585	29.01846	926.347995
2	.645243236	2586.07074	1055.75896	25.8607074	104.046925
3	.257395001	2093.82507	1046.91253	25.8496922	-25.2783681
4	.164296651	1656.58361	956.428993	25.8841189	-74.2531395
5	.0824996265	1371.40913	885.24079	27.9879415	-124.152054
6	.0496332298	1145.85021	810.238456	31.8291726	-139.868793
7	.0322838264	942.187907	719.6079	37.6875163	-142.896767
8	.0184847252	785.713666	641.532522	49.1071041	-142.651428
9	7.46227036E-03	747.08693	646.996261	83.0096589	-130.554261
10	6.4384149E-03	703.037103	641.782134	175.759276	-95.5876152
11	5.62600149E-03	623.839472	597.280821	623.839472	-62.1034591
12	3.65314963E-03	----------	----------	----------	-31

Site 1

Table VI

CALCULATIONS BASED UPON 2 FACTORS USING (COLUMN) CORRELATIO

RESULTS OF UNIQUENESS TESTS:

ROW	SPOIL
1	31.1553821
2	34.3580571
3	14.9600571
4	35.1993371
5	18.7884719
6	18.5766951
7	14.0134019
8	18.0738346
9	100.978647
10	94.5403378
11	79.9901287
12	78.5494177
13	33.0475702
14	33.8286109
15	5.23108652
16	5.31151668
17	55.2129686
18	35.570297
19	77.5514721
20	72.1592804
21	29.229459
22	28.2577331
23	99.0442893
24	82.9264476
25	20.3995395
26	19.7087812
27	43.5581192
28	39.4247272
29	10.7207413
30	10.4828033
31	104.851737
32	78.5213378
33	39.1687079
34	45.2279665
35	82.8188829
36	68.9104393
37	33.759126
38	32.4999995
39	59.015824
40	96.904984
41	31.104616
42	29.9771587

COLUMN	SPOIL
1	14.3892919
2	70.4723507
3	15.7[illegible]
4	13.3[illegible]
5	77.[illegible]
6	58.9056574
7	32.0459043
8	7.97052511
9	15.8993165
10	23.0326821
11	18.150747
12	72.011069

Table VII. Cumulative percentage contribution of eigenvalues for each eigenvector.

EIGENVECTOR	SITE 1	SITE 3	SITE 4	SITE 2
1	89.4	95.1	90.2	97.6
2	94.8	98.4	95.4	98.5
3	96.9	99.0	97.6	99.1
4	98.3	99.5	98.5	99.4
5	99.0	99.7	99.2	99.6

failed to become positive after over 20 iterations. The predicted target from the 1 factor solution when used as the target in a two factor analysis did not change on running TFA. It seems therefore that despite not being produced by the two factor analysis, the one factor solution is an acceptable target vector in a 2 factor solution. All four of the predicted 2 factor vectors produced the same vector when iterated in a 1 factor TFA.

Faced with contradictory evidence from the eigenvalue analysis it is not easy to state with any degree of certainty the number of factors determining the data set at site 1. However, TFA was run with 1 factor specified to obtain the reproduced data matrix. The error in reproduction is shown in Table VIII. TFA was then run specifying 2 factors using varying combinations of the targets produced above. All combinations produced the same error in reproduction (Table VIII) but with different factor loadings. Specifying 3 factors improved the overall R.M.S. error in reproduction but inspection of the error in reproduced data columns showed that the improvement in overall reproduction took place by producing a small percentage improvment in some columns to the considerable detriment of other columns (Table VIII).

This would seem to suggest that 2 factors describe the data most satisfactorily although the eigenvalue analysis shows that there is only one major factor in operation. A second factor is probably present but its contribution is of a similar order to that of the error in the data and consequently the program is unable to resolve it.

Table VIII

Site 1

RMS ERROR IN DATA REPRODUCTION USING 1 BASIC FACTORS = 1999.65752

RMS ERRORS FOR TFA REPRODUCED DATA COLUMNS

COLUMN	RMS ERROR
1	2180.12098
2	2167.91153
3	8.95336652
4	8.49250894
5	19.221
6	4105.67967
7	35.7345511
8	312.042958
9	4557.38346
10	19.7219802
11	527.653769
12	3.22026131

RMS ERROR IN DATA REPRODUCTION USING 2 BASIC FACTORS = 1440.75846

RMS ERRORS FOR TFA REPRODUCED DATA COLUMNS

COLUMN	RMS ERROR
1	732.097387
2	2155.3283
3	4.58114033
4	5.65342832
5	19.188674
6	3807.77117
7	25.9920393
8	88.5432458
9	2275.02581
10	12.9956777
11	209.840626
12	3.20742704

RMS ERROR IN DATA REPRODUCTION USING 3 BASIC FACTORS = 1403.43699

RMS ERRORS FOR TFA REPRODUCED DATA COLUMNS

COLUMN	RMS ERROR
1	731.929375
2	2116.17239
3	5.6467744
4	12.5194512
5	40.6905816
6	3631.949
7	142.271848
8	429.023941
9	2272.09238
10	13.8942524
11	257.182449
12	3.08796129

It is reasonable to postulate the predicted vector of the one factor solution as the profile for the major factor at site 1 and that this may match the immediate local geology. A search of the chemical and geological literature has however failed to find a chemical profile of millstone grit which could be compared to this target, or which could itself be used as a target. This appears to be due to millstone grit rarely producing any important mineralisation and hence its chemical composition appears to be of little interest to geologists.

Site 3.

The initial EVA results are shown in Table IX. The PIND, IND and other functions suggest that 4 factors adequately determine the factor space. After performing uniqueness tests 3 rows were removed from the data matrix and EVA repeated once again (Table X). The number of factors as indicated by the eigenvalues and the IND function now reduced to 2. The uniqueness test shows a low SPOIL for column 3 (Table XI) and removal of this column was considered. In order to maintain comparisons with the other sites however it was decided to leave this column in the data matrix. The cumulative percentage contribution of the eigenvalues is shown in Table VII. It can be seen that for this site 1 factor describes 95% of the variation with 2 factors describing 98.3%.

Iterating target vectors of unity and zero as was done

Table IX

REFERENCE: CYNON SEDIMENT DATA:SITE 3

SUMMARY OF EVA RESULTS USING (COLUMN) CORRELATION

NF	EV	RE	IE	IND	PIND
1	11.3661261	3413.7128	985.454003	28.2125025	768.851625
2	.336142073	2453.77367	1001.74891	24.5377367	137.744738
3	.123504335	1978.60424	989.30212	24.4272128	59.4273034
4	.082547207	1522.35094	878.929728	23.7867335	34.0411985
5	.0325059161	1307.49349	843.98342	26.6835406	-.0694282055
6	.0257427704	1061.51114	750.601726	29.4864206	1.91905069
7	.0162725043	833.073866	636.270675	33.3229546	-4.9921855
8	9.70189397E-03	614.014316	501.34059	38.3758948	-13.0773414
9	4.12980584E-03	473.59867	410.148479	52.6220744	-24.314122
10	2.29282806E-03	323.427253	295.247337	80.8568133	-22.8538549
11	7.18181133E-04	252.929237	242.161308	252.929237	-24.2382588
12	3.16340109E-04	----------	----------	----------	-13

Table X

REFERENCE: CYNON SEDIMENT DATA:SITE 3

SUMMARY OF EVA RESULTS USING (COLUMN) CORRELATION

NF	EV	RE	IE	IND	PIND
1	11.4179087	3243.61638	936.351396	26.8067469	733.274635
2	.391070507	1948.81478	795.600301	19.4881478	167.393168
3	.0690439237	1641.52576	820.762879	20.2657501	33.6092891
4	.0558864239	1281.60591	739.935515	20.0250923	25.5339232
5	.0231692876	1104.12127	712.707215	22.5330871	.664304733
6	.0209038815	854.153759	603.977915	23.7264933	3.55126187
7	9.97906293E-03	691.871701	528.425741	27.6748681	-11.5256384
8	5.21941248E-03	582.174256	475.34329	36.385891	-13.6593151
9	4.42783578E-03	398.063696	344.733273	44.2292996	-7.80542827
10	1.50228677E-03	297.219713	271.323236	74.3049283	-18.3338081
11	6.85389074E-04	201.02214	192.464046	201.02214	-15.0266255
12	2.03248434E-04	----------	----------	----------	-10

Table XI

CYNON SEDIMENT DATA:SITE 3

CALCULATIONS BASED UPON 2 FACTORS USING (COLUMN) CORRELATIO

RESULTS OF UNIQUENESS TESTS:

ROW	SPOIL
1	13.4025806
2	12.1865472
3	13.9507869
4	14.84124
5	50.0302853
6	11.7572378
7	49.4270815
8	69.390177
9	64.3049969
10	17.9689878
11	17.1169102
12	80.3255235
13	102.474858
14	91.7796228
15	102.189701
16	80.0944837
17	63.093474
18	17.3559751
19	18.0317635
20	11.8330351
21	11.4415405

COLUMN	SPOIL
1	29.4892549
2	43.6928417
3	2.99312794
4	74.0688286
5	25.3777935
6	25.4050896
7	22.0050175
8	43.8448577
9	28.3645842
10	82.2603828
11	59.0471962
12	58.4534897

for site 1 produced a similar pattern of predicted vectors i.e. a single predicted vector from all targets if 1 factor was specified in the TFA while 5 predicted vectors were produced if 2 factors were specified.

Site 3 would seem to have 2 factors contributing to it, a major one and a second factor, which, as for site 1 the TFA program seems unable to resolve, probably due to error in analytical data.

Site 4.

Two rows of the original data matrix were removed due to incomplete data on those samples. The eigenvalue analysis and results of the uniqueness test are shown in Tables XII and XIII. There being no outstandingly low values no further data was removed from the matrix.

The cumulative percentage contribution of each factor is shown in Table VII. Once again, there is obviously 1 major factor operating, together with 1 or more other factors. The TFA, as for the previous 2 sites produced 1 vector for a 1 factor solution but 6 vectors for a 2 factor solution. The major factor describes some 90% of the data, the second factor an additional 5%.

Site 2.

The eigenvalue analysis and uniqueness test

Table XII

Site 4

SUMMARY OF EVA RESULTS USING (COLUMN) CORRELATION

NF	EV	RE	IE	IND	PIND
1	10.8265031	3475.84787	1003.39085	28.7260155	576.6411
2	.624428467	2493.61476	1018.01396	24.9361476	142.706475
3	.266805055	1884.61274	942.306372	23.266824	71.3291086
4	.110352019	1559.99565	900.663911	24.3749321	27.6834331
5	.0804997241	1216.09577	724.986446	24.8182811	17.5854711
6	.0515388788	867.517945	613.427822	24.0977207	-2.89650604
7	.0138865668	767.18865	585.950011	30.687546	-36.1828345
8	.0129271291	608.0532	496.473359	38.003325	-25.515795
9	8.07407234E-03	433.796646	375.678915	48.1996273	-26.435571
10	2.22595579E-03	395.254693	360.816519	98.8136732	-32.969147
11	1.97649972E-03	297.686098	285.01274	297.686098	-19.645116
12	7.82500884E-04	----------	----------	----------	-11

Table XIII

Site 4

CALCULATIONS BASED UPON 3 FACTORS USING (COLUMN) CORRELATION

RESULTS OF UNIQUENESS TESTS:

ROW	SPOIL
1	10.7921614
2	9.59934918
3	8.34420323
4	9.14691098
5	10.4526629
6	10.2997684
7	35.8389844
8	21.742046
9	6.36795295
10	5.99603776
11	14.6532079
12	33.8444712
13	5.74837171
14	7.9092205
15	19.6745865
16	18.0574116
17	7.51948999
18	6.31398325
19	38.3782596
20	33.8560769
21	79.0505713
22	69.1388618

COLUMN	SPOIL
1	16.8338443
2	8.99202535
3	5.52100685
4	11.2189402
5	66.4128758
6	13.1453227
7	3.58210688
8	39.8412718
9	5.73253633
10	43.4960607
11	30.5695772
12	14.5115847

highlighted two rows of the data matrix with particularly low SPOIL values. These were removed from the matrix and eigenvalue analysis and uniqueness test was repeated. This disclosed a further two rows which were removed and the analysis repeated. The results are shown in Tables XIV and XV. These are significantly different from the results so far obtained in that the uniqueness test results are all high and show a remarkable consistency not observed for any of the other sites.

All the indicators of the eigenvalue analysis pointed to 1 factor only. The cumulative percentages of the eigenvalues (Table VII) show that 97.6% of the variation can be accounted for by 1 factor. The target vectors of unities and zeros rapidly converged to the same predicted vector. This predicted vector along with the other three produced for the other sites is shown in Table XVI.

From these results it is clear that there is one major factor controlling the heavy metal concentration at each site. For sites 1,3 and 4 a second factor may also be detected. A second factor appears to be making a significant contribution at site 1 (5.4%) and similarly at site 4 (5.2%), but is less important at site 3 (3.3%). All the indicators show that the profile at site 2 is governed overwhelmingly by 1 factor.

The predicted vectors obtained by applying a 1 factor solution in the target vector analysis (Table XVI) must be prime candidates as profiles for the major factor operating at each site. Two of these profiles, those at

Table XIV

Site 2

SUMMARY OF EVA RESULTS USING (COLUMN) CORRELATION

NF	EV	RE	IE	IND	PIND
1	11.7093549	2345.31891	677.035252	19.3828009	1520.4275
2	.108269171	1948.50038	795.47195	19.4850038	-37.4438757
3	.0674045756	1630.7609	815.380449	20.1328506	-83.2068763
4	.0356611172	1436.60156	829.422298	22.4468994	-113.778388
5	.0297262309	1214.33484	783.849769	24.7823437	-107.849767
6	.0208791468	997.972764	705.673309	27.7214657	-114.817738
7	.0140443568	781.278508	596.711317	31.2511403	-124.587527
8	4.89246956E-03	713.000541	582.162505	44.5625338	-140.541753
9	4.24367687E-03	619.150383	536.199961	68.794487	-108.985267
10	3.41409258E-03	468.670871	427.836013	117.167718	-82.6355423
11	1.21651347E-03	431.339999	412.976608	431.339999	-66.9819731
12	8.93729193E-04	----------	----------	----------	-33

Table XV

Site 2

CALCULATIONS BASED UPON 1 FACTORS USING (COLUMN) CORRELATIO

RESULTS OF UNIQUENESS TESTS:

ROW	SPOIL
1	153.082276
2	154.162115
3	157.798794
4	156.998649
5	116.132495
6	112.805828
7	132.976467
8	134.706201
9	144.754384
10	140.059416
11	149.027021
12	147.945688
13	137.53847
14	138.334236
15	139.906713
16	143.514385
17	143.277006
18	146.431193
19	160.822936
20	160.786643
21	139.512284
22	150.614098
23	134.295032
24	136.06003
25	126.647199
26	126.231504
27	188.81193
28	178.147287
29	134.232677
30	134.971272
31	117.512992
32	118.694678
33	126.183982
34	129.303988
35	128.489137
36	132.621146
37	167.161301
38	155.70916
39	127.403511
40	129.682455
41	133.954464
42	130.497691
43	131.479822
44	128.220052

COLUMN	SPOIL
1	134.243758
2	132.701226
3	133.479341
4	134.372496
5	132.871385
6	134.725806
7	132.433898
8	132.507289
9	133.4683
10	132.46595
11	133.002447
12	138.306736

Table XVI

Predicted target vectors produced by TFA for each site.

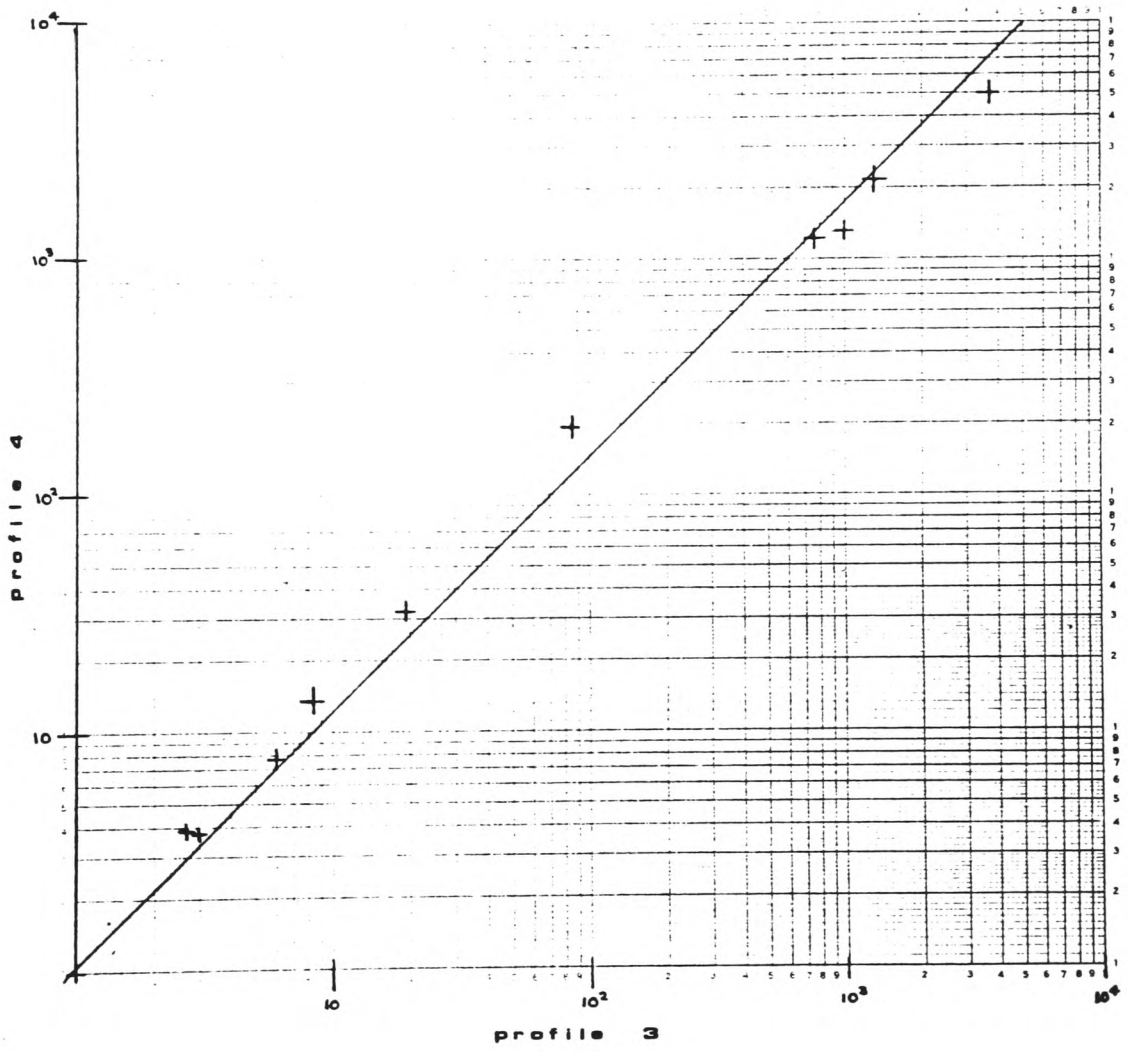
	Site 1	Site 3	Site 4	Site 2
Al	359	785	1212	429
Ca	389	1309	2169	597
Cd	1.5	0.08	0.3	0.2
Cr	1.0	1.0	1.0	1.0
Cu	2.0	6.0	7.7	7.4
Fe	1625	3712	5075	4131
Pb	10.0	8.4	13.3	9.2
Mg	30.5	2.7	3.9	197
Mn	229	86	194	61
Ni	4.2	3.0	3.7	3.7
Zn	106	19.4	32	13.7

sites 3 and 4 appeared similar. They were therefore plotted against each other, Fig 13. It can be seen that the points lie very close to a straight line indicating the similarity in the profiles. (correlation coefficient r = 0.997)

Following this observation correlations between the other sites were checked for. This revealed a good correlation between the site 2 and site 3 profiles (r = 0.97).

The major factor at site 1 is most likely to be the inherent geology of the site with the minor factors due to additional sources. The similar profiles generated for sites 2, 3 and 4 suggest a common source.

Figure 13. Profile 3 v Profile 4.



Pattern Recognition

The purpose of the techniques available in pattern recognition is to determine whether any patterns exist in the data and if possible to extract information about those patterns. For example, it would be of interest to discover whether samples from a particular site can be distinguished from those taken from other sites in the river. The techniques used in this study were all contained in a computer package - ARTHUR (68). The methods used fall into two classes, supervised and unsupervised learning methods.

Supervised learning methods use data which is categorised. Models are then formed which attempt to distinguish between the categories (in this study, the individual sites). Unsupervised learning methods do not require foreknowledge of any categories and it is then the responsibility of the observer to determine whether the patterns revealed by these methods coincide with any preconceived categories.

Hierarchical Cluster Analysis

This is one particular example of a range of unsupervised learning methods collectively called cluster analysis.

Each data point (sample) is described by a number of measurements (features) which in this case are the heavy metal concentrations. The data points can be plotted as n variables in n dimensional space. This is difficult to

visualise but is only a mathematical extension of plotting 3 variables in 3 dimensional space. The points which lie closest together in this n-dimensional space are said to be similar.

In hierarchical clustering the euclidean distances are first calculated to provide the distance matrix. The euclidian distance may be described as the shortest distance between two points. If the points were plotted in two dimensions the distance may be calculated as the length of the hypotenuse of a right angle triangle for which the points form its apices. This may be extended to n-dimensions using the following equation.

$$D_{kl} = \sum_{j=1}^{n} (x_{kj} - x_{lj})^2$$

The two closest points are merged to form a new point so reducing the number of data points by one. A new distance matrix is calculated and the process repeated. To prevent the eventual compression of the data to just one point a distance is specified above which two points are not merged. This will result in clusters of similar samples (close points). There are several ways of merging elements to form new clusters. In ARTHUR the complete linkage method was used in which the distance between two clusters is considered to be equal to the largest distance between two individual elements, one of each cluster.The results are plotted as a dendogram. Fig. 14.

Figure 14. Cluster Analysis Dendogram

D
E

HIERARCHICAL DENDROGRAM PLOT -- PAGE 1

COMPLETE LINK METHOD OF CLUSTERING

S I M I L A R I T Y M A T C H I N G

133	89		6	1.00	**
134	89		6	1.00	**
129	80		8	1.00	********
135	90		7	1.00	** **
136	91		7	1.00	** **
141	96	1	0	1.00	**** **
142	97	1	0	1.00	**** *********
143	100	1	2	1.00	**** * *
144	101	1	2	1.00	**** ******* *
103	55		3	3.00	**** * *
104	56		3	3.00	***** * *
121	73	1	2	3.00	***** ** * *
122	74	1	2	3.00	**** ** ********** *
130	85		8	1.00	******* *****
125	80		2	1.00	*** * *
126	81		2	1.00	****** * *
138	93		8	1.00	**** *** * *
127	82		3	1.00	*** ** * * *
128	83		3	1.00	*** ********** *
139	94		8	1.00	**** * *
140	95		8	1.00	******** *
137	92		8	1.00	***** ********
99	51		1	3.00	***** * *
100	52		1	3.00	***** ********** * *
117	69	1	0	3.00	**** * ******** * *
118	70	1	0	3.00	**** ********** * * *
109	61		6	3.00	**** * * *
110	62		6	3.00	***** * * *
123	78		1	1.00	*** * * * * *
124	79		1	1.00	****** *** ******* *
105	57		4	3.00	**** * * *
106	58		4	3.00	********* * * *
101	53		2	3.00	**** * * *
102	54		2	3.00	*** * * * *
119	71	1	1	3.00	*** ** * ******** *
120	72	1	1	3.00	**** * * *
115	67		9	3.00	**** ** * *
116	68		9	3.00	***** * *
111	63		7	3.00	*** * ***** * *
112	64		7	3.00	**** * * * *
107	59		5	3.00	**** * **** *
108	60		5	3.00	***** * *
113	65		9	3.00	** * *
114	66		9	3.00	********** *
95	100	2	1	2.00	**** *
96	101	2	1	2.00	**** * *
87	92	2	0	2.00	**** * * *
88	93	2	0	2.00	***** * *
89	90	2	1	2.00	**** * *
90	95	2	1	2.00	***** * * ******
97	102	2	5	2.00	* ** * * * *
98	103	2	5	2.00	***** * * * *
83	88	1	9	2.00	* ***** * * *
84	89	1	9	2.00	** * ** * *
81	86	1	7	2.00	**** ** * *
82	87	1	7	2.00	**** ** ** * *
53	56		2	2.00	* ******** * *
54	57		2	2.00	******* * * *

The data can be seen to be divided into 6 major clusters:

A This contains 4 samples from site 1 which from their similarity measure can be seen to be very different from all the other samples. The 4 samples involved are in fact identified as 'outliers' by most of the other pattern recognition techniques.

B contains 16 samples from site 1 which are in a slightly different cluster to

C which contains the remaining site 1 samples.

D This cluster contains all but two of the site 2 samples.

E This cluster contains all except 4 of the site 3 samples, plus 2 site 4 samples.

F This remaining group contains all except 4 of the site 4 samples plus 4 of the site 3 samples.

Z The samples contained in this section of the dendogram are in fact two separate, small clusters. One contains 2 site 2 samples that were not encompassed by cluster D and 2 site 4 samples that are somewhat different from cluster F. These samples were identified as 'outliers' for their sites by the other pattern recognition techniques.

It can be seen that this technique of pattern recognition is capable of dividing the data into its correct categories albeit not perfectly. (viz. the site 3 samples that appeared in cluster E). It is capable of identifying 'outliers' in the data i.e. samples that for some reason are very much different from the others in the group.

Principal Component Plots

Visualising n dimensional plots (when n>3) is a difficult if not impossible concept. The human brain is however very adept at recognising patterns in one, two or three dimensions. Principal component plots are an attempt to represent n dimensional data in two dimensions thereby simplifying the problem in pattern recognition. This reduction in dimensionality is achieved by finding an axis in the n dimensional space that accounts for the largest variation in the data. A second axis which accounts for the greatest variation in a direction orthogonal to the first is then computed. A projection of the data on to this plane then produces a two dimensional plot. Any reduction in dimensionality results in loss of information and so the plot will be less than perfect. Obviously third, fourth, ... nth axes can be found, so reducing loss of information. However since the principal components are ordered according to importance, provided that the number of components used contain a large proportion (around 90 %) of the variation then there will not be too great a loss in information.

The principal components are linear combinations of the original variables. The order of importance of the principal components is determined by the eigenvalues of the eigenvectors extracted from the covariance matrix.

Mathematically, the data matrix X may be viewed as a set of variables represented by vectors of measurements.

$$X = (\vec{x}_1, \vec{x}_2, \vec{x}_3 \ldots \vec{x}_n) \quad \text{where} \quad \vec{X}_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ x_{3j} \\ \vdots \\ x_{nj} \end{pmatrix}$$

these vectors are transformed so that the objects (samples) are represented by a new set of variables, each a linear combination of the original variables.

$$U'_j = l_j' \cdot X'$$

where $\vec{l}'_j = (a_{1j} \ldots\ldots a_{nj})$ and $X' = \begin{pmatrix} \vec{x}'_1 \\ \vec{x}'_2 \\ \vdots \\ \vec{x}'_n \end{pmatrix}$

The principal components (U s), must be uncorrelated hence the matrix equation

$$C\ L' = L'\Lambda$$

where C is the covariance matrix

L is the matrix of column vectors $\vec{l}_j$

and Λ a diagonal matrix with diagonal elements $\lambda_1 \ldots \lambda_n$

the l s are the eigenvectors with elements a and the λ s are the eigenvalues.

The covariance matrix is obtained by multiplying the m x n data matrix $X_{(mxn)}$ by its transpose $X'_{(nxm)}$ and subtracting the mean of each variable to obtain an n x n matrix. Dividing the elements by the number of objects minus 1 gives the covariance matrix C.

Fig.15 shows the plot for principal component 1 v principal component 2. Similarly Fig.16 shows principal component 3 v principal component 4. These plots demonstrate once again that the samples can be classified according to their sites. The results are similar to the hierarchical clustering, in that site 1 is shown to contain more than one subset, site 2 shows a fairly compact group. The plot as with hierarchical clustering does not resolve sites 3 and 4 quite as successfully as it does sites 1 and 2. It is also interesting to note that principal component 1 is constant for sites 3 and 4 and the resolution of these sites (such as it is) depends solely on principal component 2. This fact supports the factor analysis which indicated that the major factor at sites 3 and 4 were essentially the same profile. A further interesting observation is that the plot for components 3 v 4 produce a wider separation than that for component 1 v 2.

Figure 15. Principal component plot 1 v 2

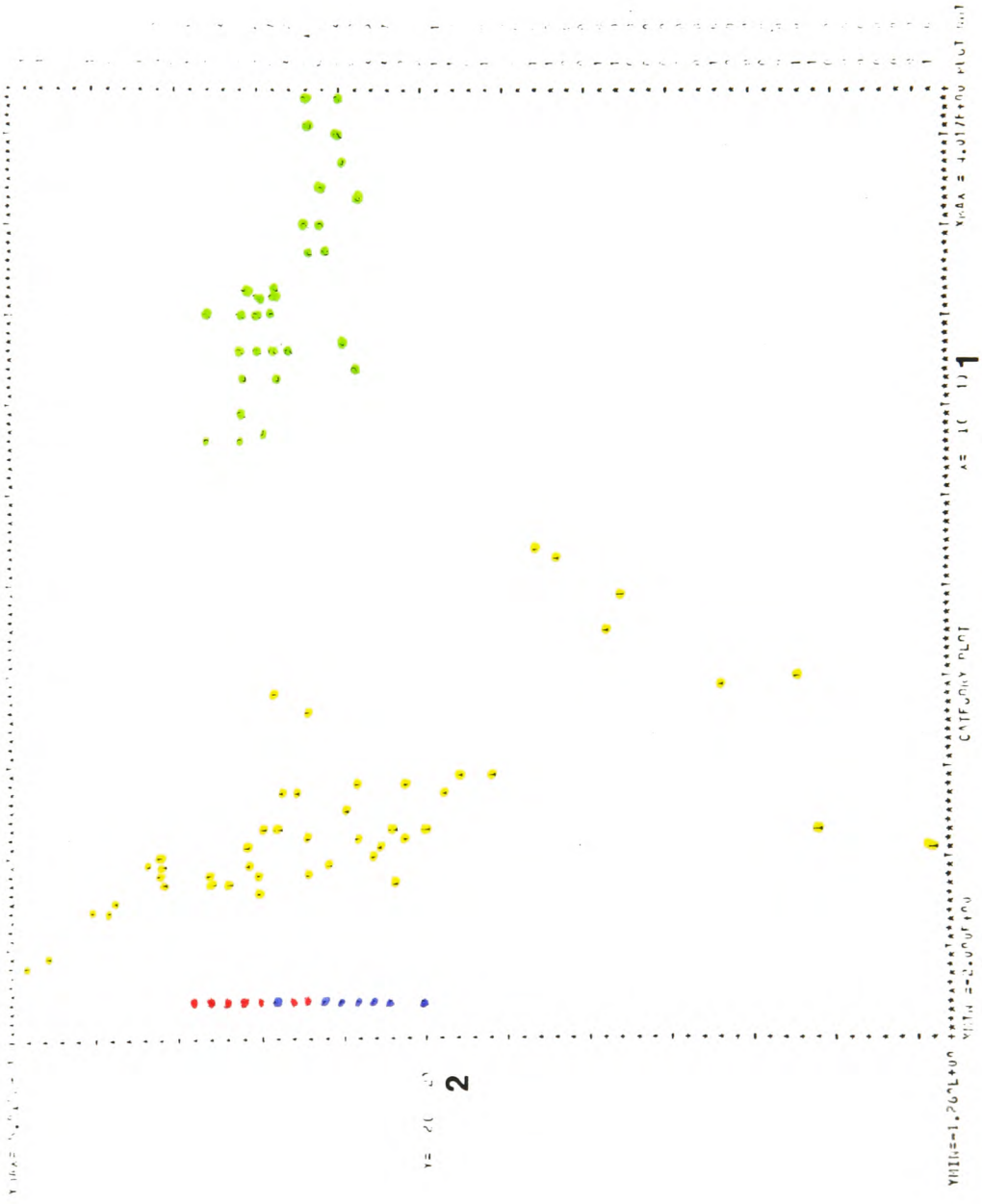
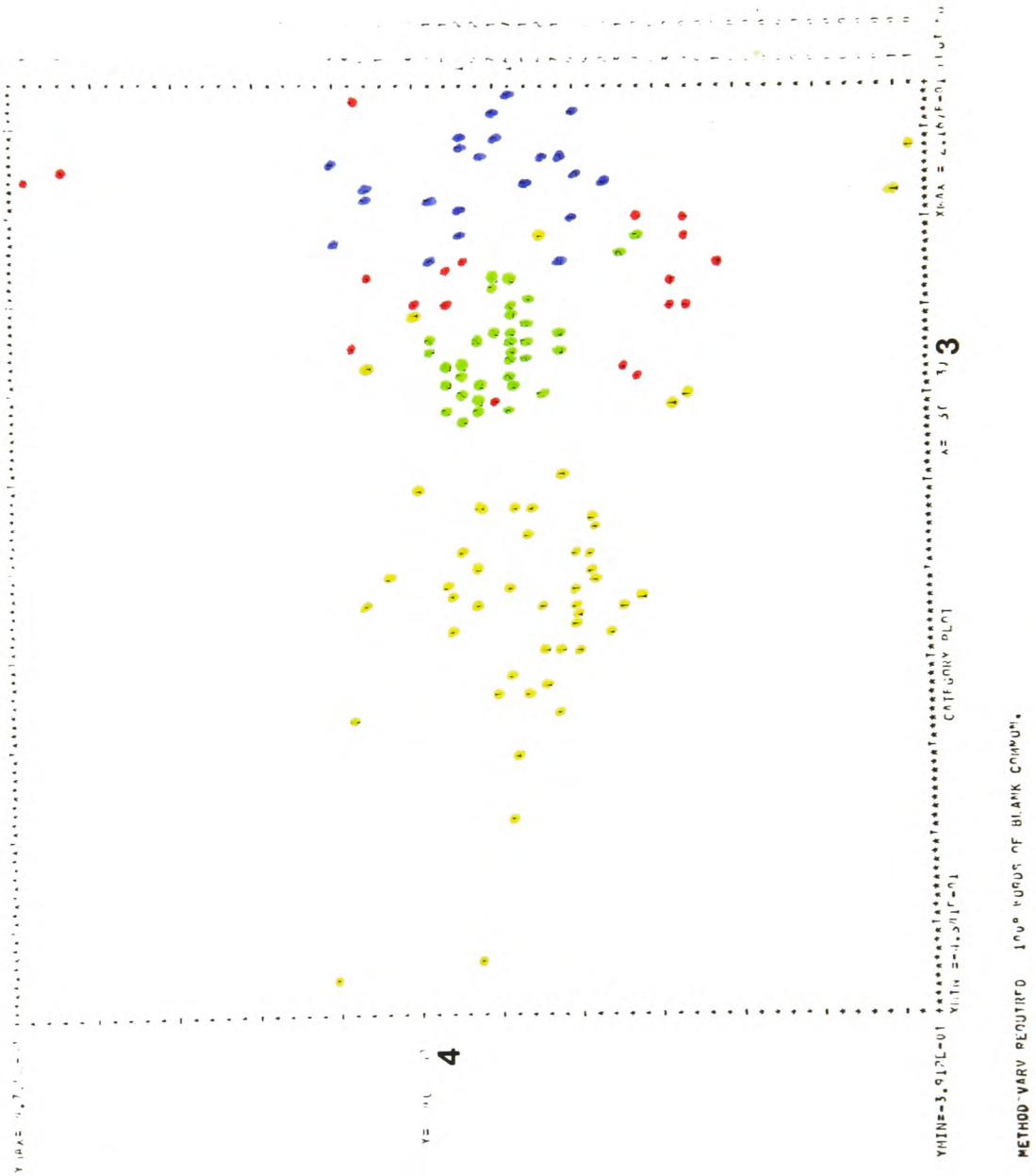


Figure 16. Principal component plot 3 v 4.



Non-linear mapping.

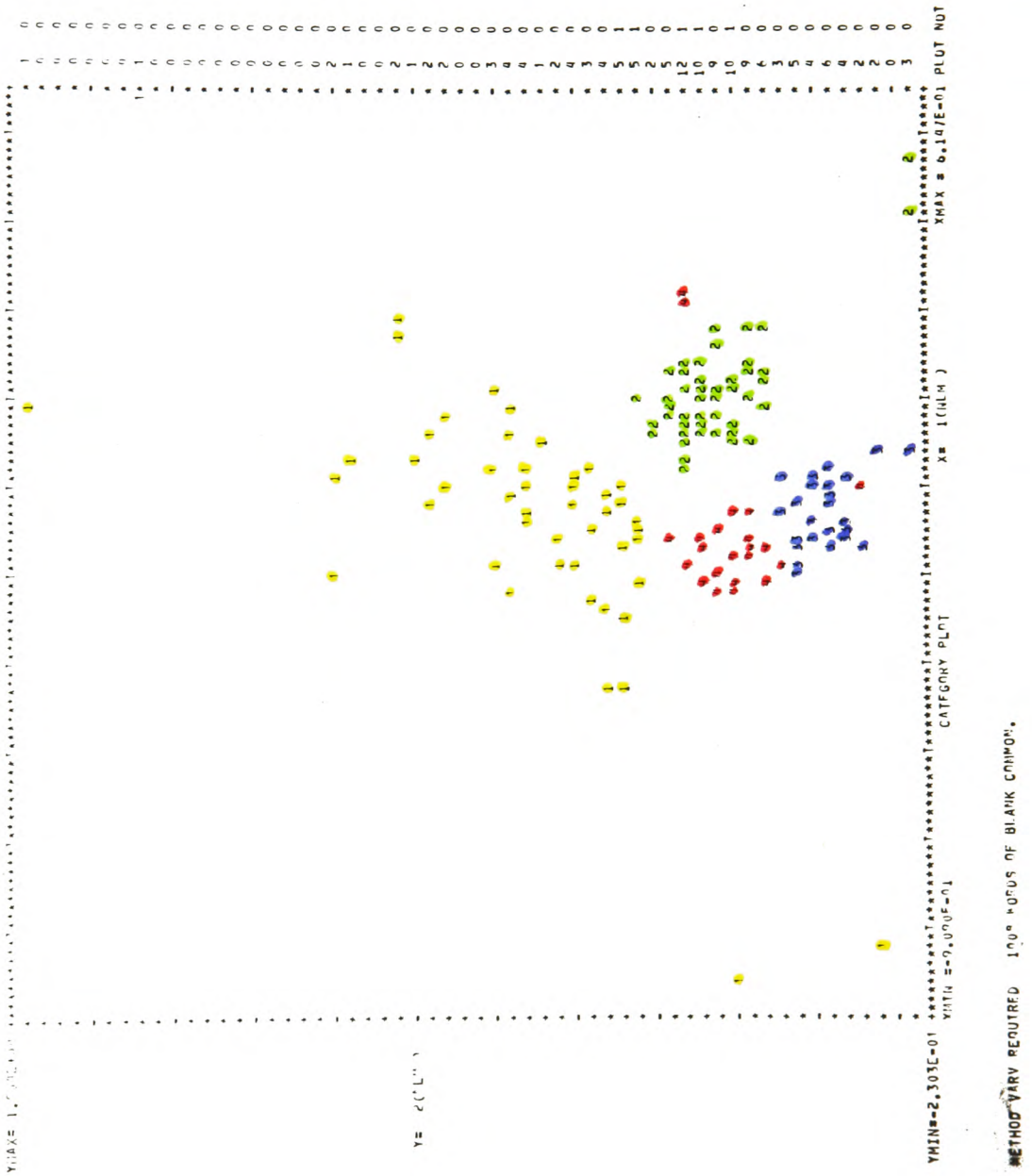
This is another method of representing an n dimensional space in 2 dimensions. As with the principal component plots this is bound to result in some error and loss of information. Nevertheless, it can still provide a useful means of detecting patterns in the data. The basis of non-linear mapping is that the 2 dimensional representation of the data points should reflect as accurately as possible the relationship between the points in n-dimensional space i.e. the distance between points in the 2 dimensional space should equal the distance between points in the n-dimensional space. This is not possible without introducing error, so an optimum representation is produced by minimising the error between distances in the 2 dimensional space compared to the n-dimensional space. This is achieved by an iterative process.

The method was first used in analytical chemistry by Kowalski and Bender (69). The original method proposed by Sammon (70) was modified by Kowalski and Bender (71) for ARTHUR.

The results are shown in Fig.17. The method has once again successfully distinguished the four sites. It has also detected 'outliers'.

Comparing the results of the non-linear mapping with hierarchical clustering reveals several similarities. The outliers in the non-linear mapping (41,42,46,47 site 1; 84,85 site 2; 85,86,87 site 4) reflect the clusters found in the hierarchical clustering. Group A found the

Figure 17. Non linear mapping.



outliers of site 1, Group Z contained the site 2 outliers and two (86,87) of the site 4 outliers. The third site 4 outlier (85) was placed in the small cluster of site 3 samples found in group F of the hierarchical clustering. The resolution of sites three and four by this technique shows a significant improvement over the other methods.

Feature - feature plots.

In this study the features are the heavy metal concentrations. The features are weighted on the basis of their individual value in describing the data matrix. Several methods of weighting the data are available. Examples of those in common use are variance and Fisher weights. Variance weighting assumes that features with the greatest variance are more important to classification than those with a small variance. A weight is calculated for each feature which is the ratio of "interclass variability" to "intraclass variability".

Fisher weights are calculate differently, the Fisher ratio given by the equation

$$F = \frac{(m_{i1} - m_{i2})^2}{v_{i1} + v_{i2}}$$

m = mean

v = variance

for feature i classes 1,2

In this study Fisher weights were used. The features to be plotted were selected using the ARTHUR SELECT routine. This selects the feature with the highest weighting and

decorrelates the remaining features. These are reweighted and the feature with the new highest weight is selected. This process continues until a specified number of features is chosen or a specified minimum weight attained.

Figs. 18,19 show the plots of Mg v Zn and Mg v Fe respectively. It can be seen that sites 1 and 2 are separated quite successfully but differentiation between sites 3 and 4 is very poor. Furthermore these plots showed a similarity to the principal component plots leading to the conclusion that the first principal component was heavily dependent upon the magnesium concentrations. These concentrations were therefore removed from the data matrix and the principal component analysis and feature-feature plots were repeated.

No real improvement resulted leading to a reappraisal of the suggestion that it is the magnesium concentration that determines the separation. A revised hypothesis was made that suggested that sites 3 and 4 are so similar that **any** major features or principal component that discriminates between sites 1 and 2, inevitably results in a poor classification of sites 3 and 4. This then led to the conclusion that principal components lower in the order may more successfully discriminate between sites 3 and 4. Despite imperfect resolution this was confirmed by principal component plot 3,4 (fig. 16) which contained the magnesium concentrations. Similar plots without magnesium improved the resolution even further (figs 20,21).

The feature-feature plots reflected the principal component plots in that the major selected features only

Figure 18. Feature-feature plot Mg v Zn.

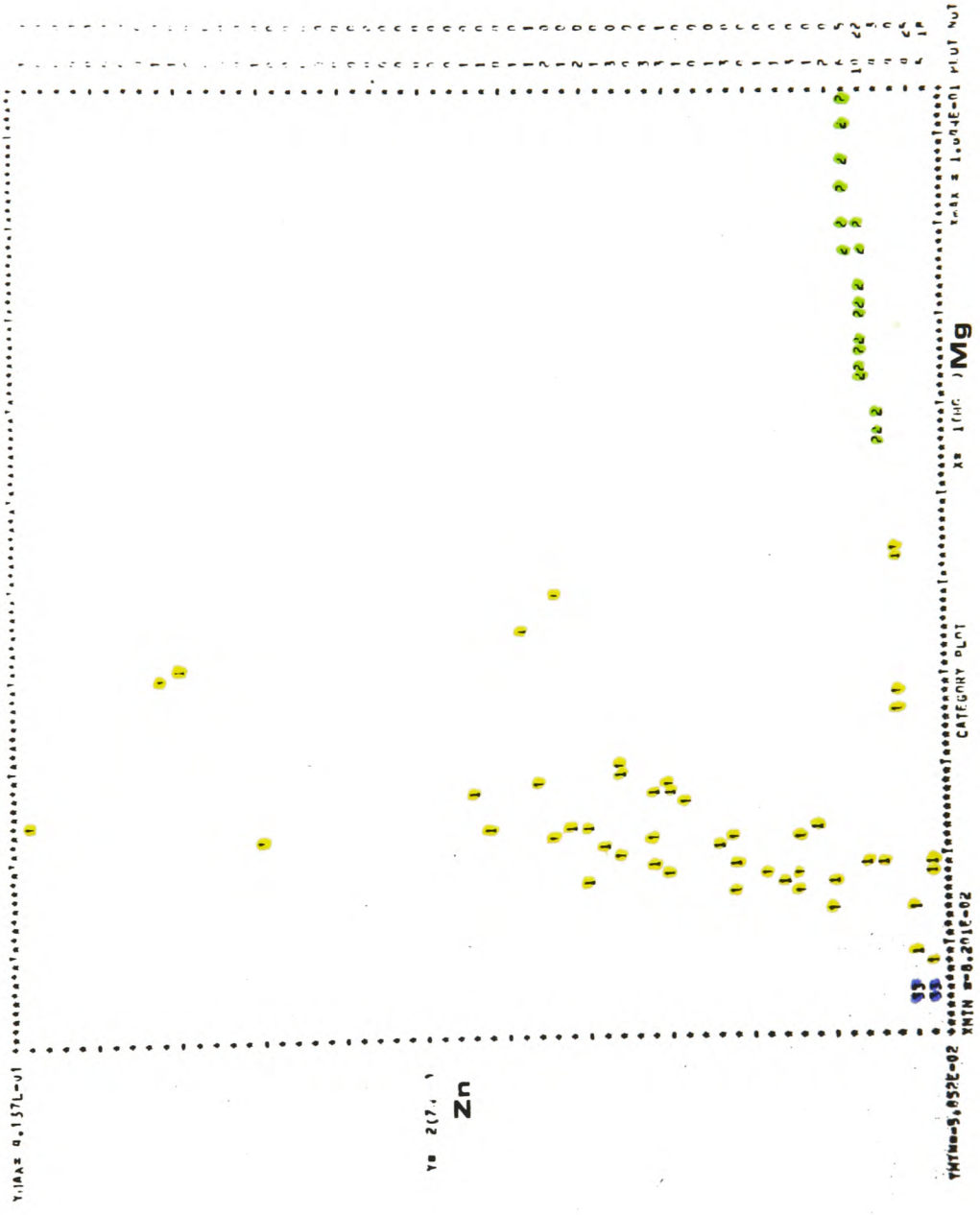


Figure 19. Feature-feature plot Mg v Fe.

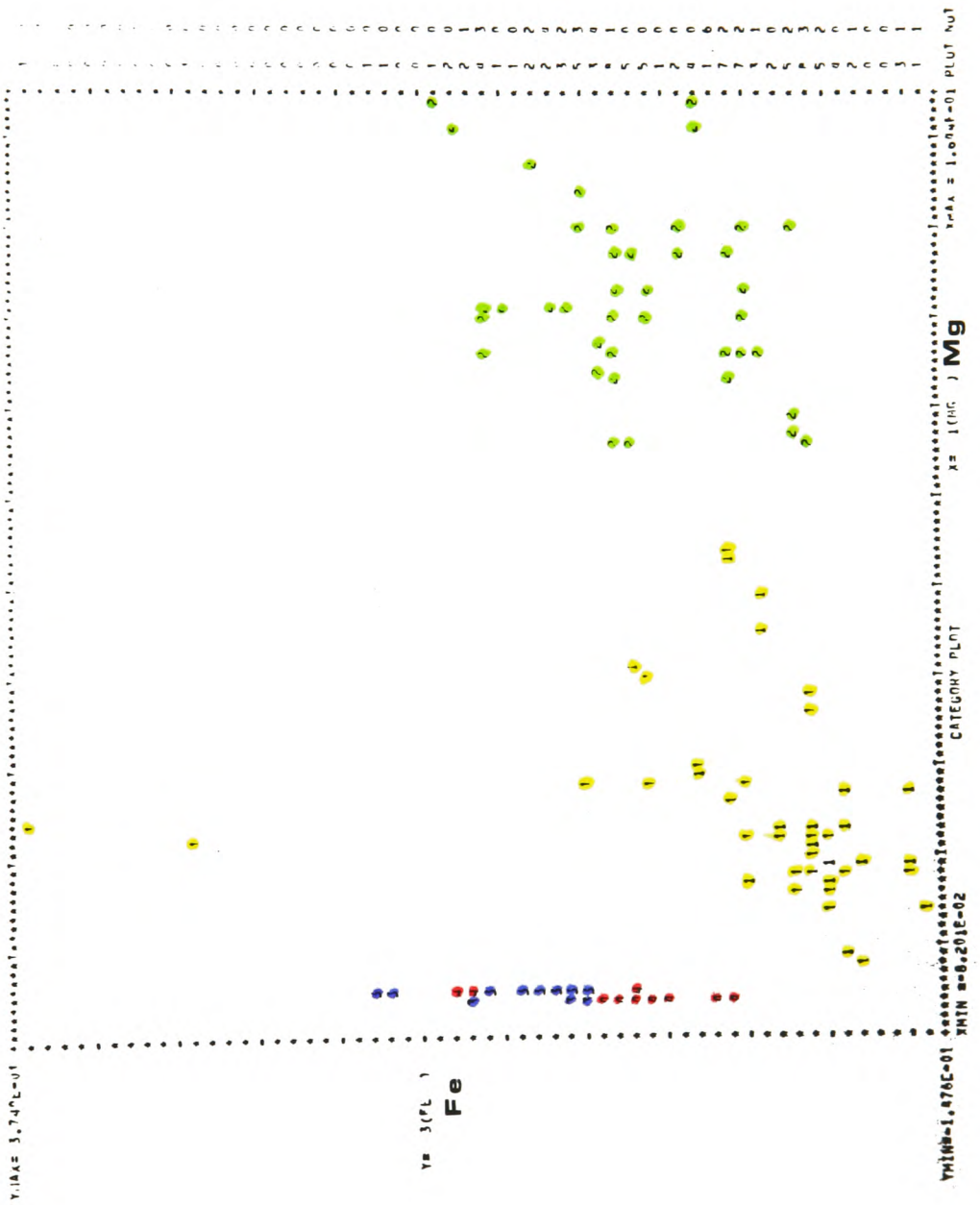


Figure 20. Principal component plot 1 v 3 after removing Mg concentrations.

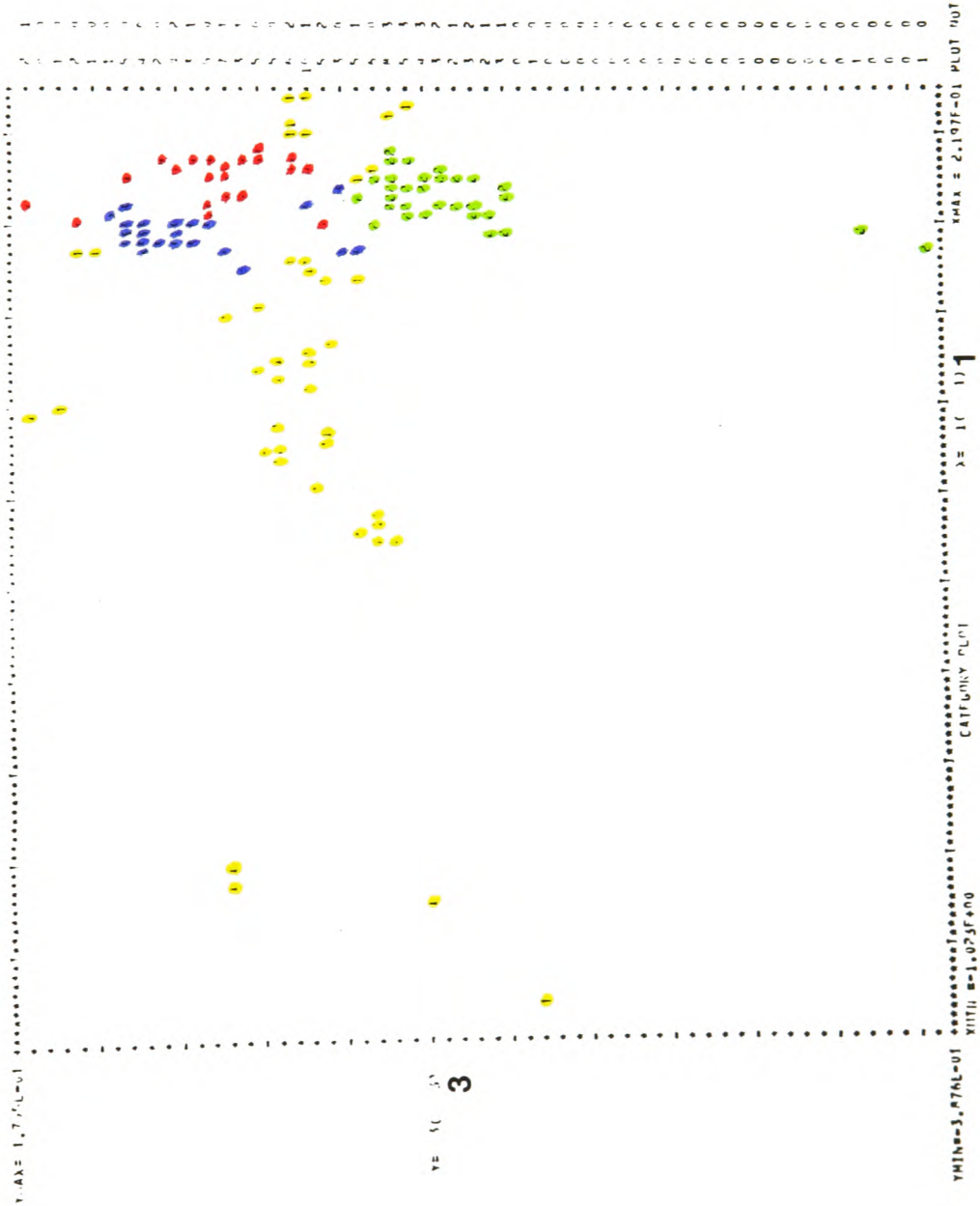
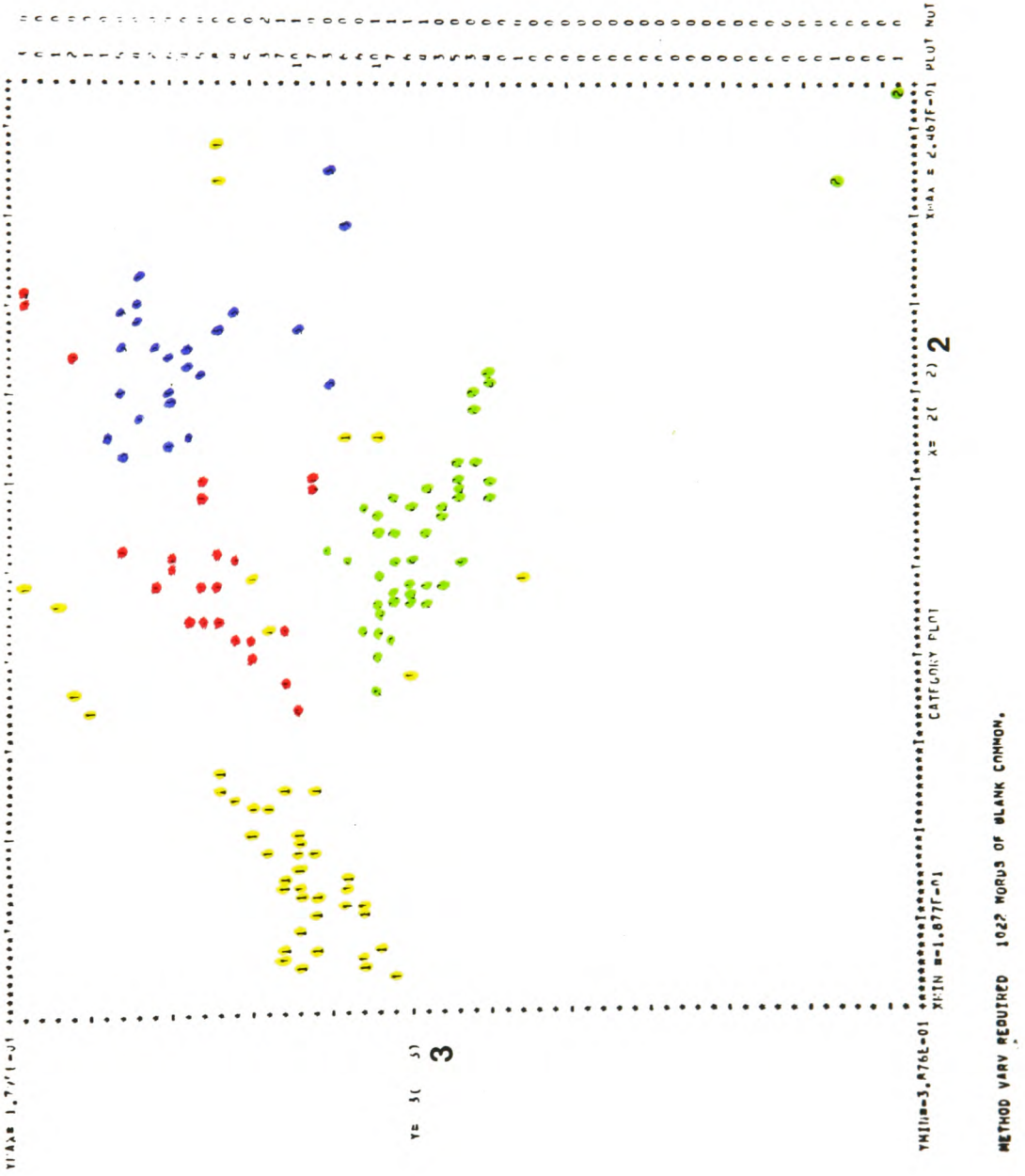


Figure 21. Principal component plot 2 v 3 after removing Mg concentrations.



poorly resolved sites 3 and 4 whereas features less important to the overall classification begin to discriminate between them (fig 22).

It was realised that the presence of sites that are considerably different makes resolution of similar sites difficult and that resolution of these sites ought to be achievable by removing these sites (categories). To confirm this hypothesis sites 1 and 2 were removed from the data matrix. The principle component plot (1 v 2) for sites 3 and 4 only are shown in figure 23 (ARTHUR has renumbered them 1 and 2). This plot shows that the sites can be resolved.

Supervised learning

Only one supervised learning technique was used to determine whether the sites could be correctly classified.

k-Nearest-Neighbour (kNN)

The simplest kNN method with k = 1 classifies a test object into the class which is the closest to the test object. If k > 1 the test object is classified into the class to which the majority of the k closest members belong. This is called the 'majority vote' or 'committee vote' procedure.

A disadvantage of this method is that all points are classified regardless of the actual distances involved and so outliers are not identified. A second disadvantage was

Figure 22. Feature-feature plot Al v Ni.

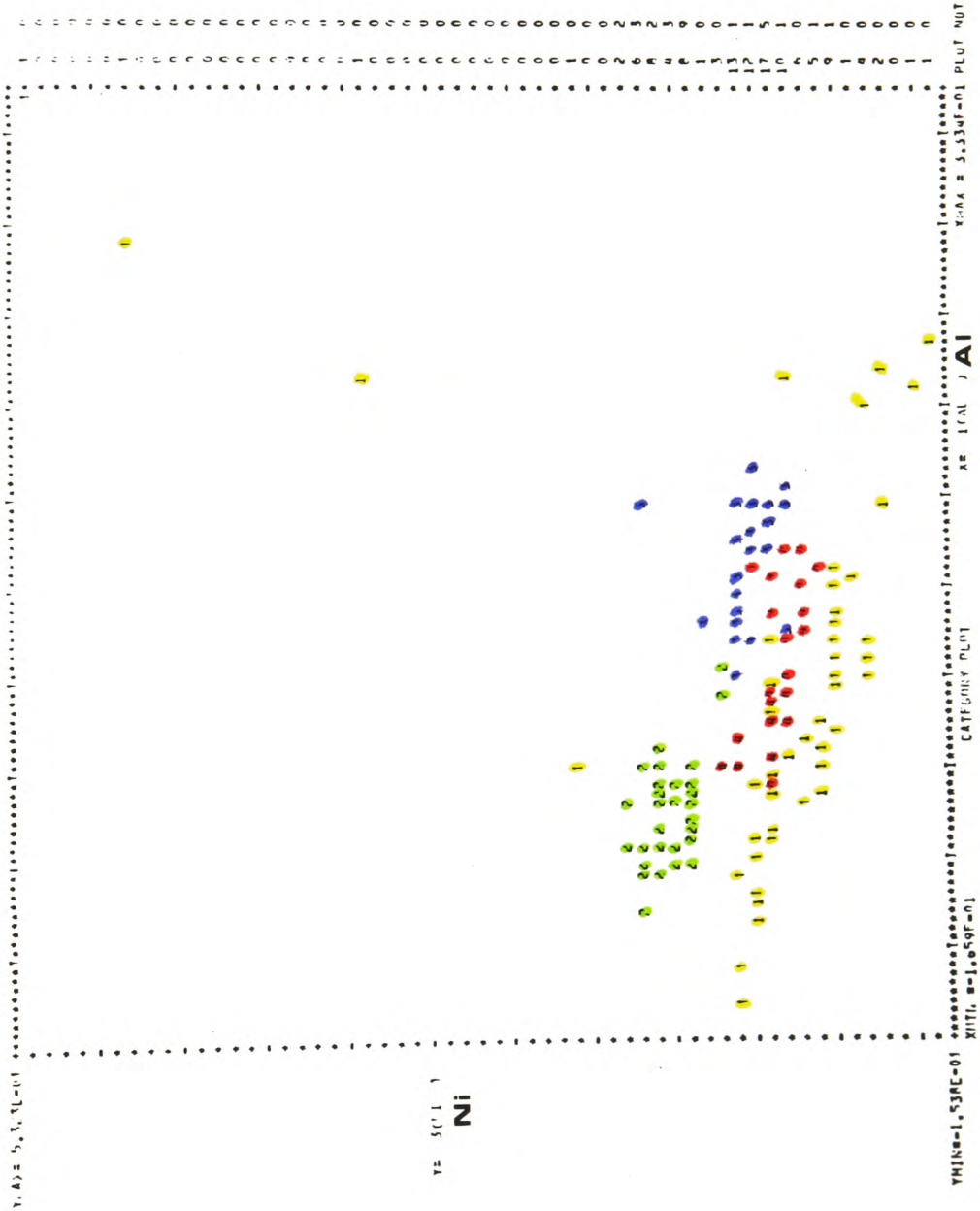
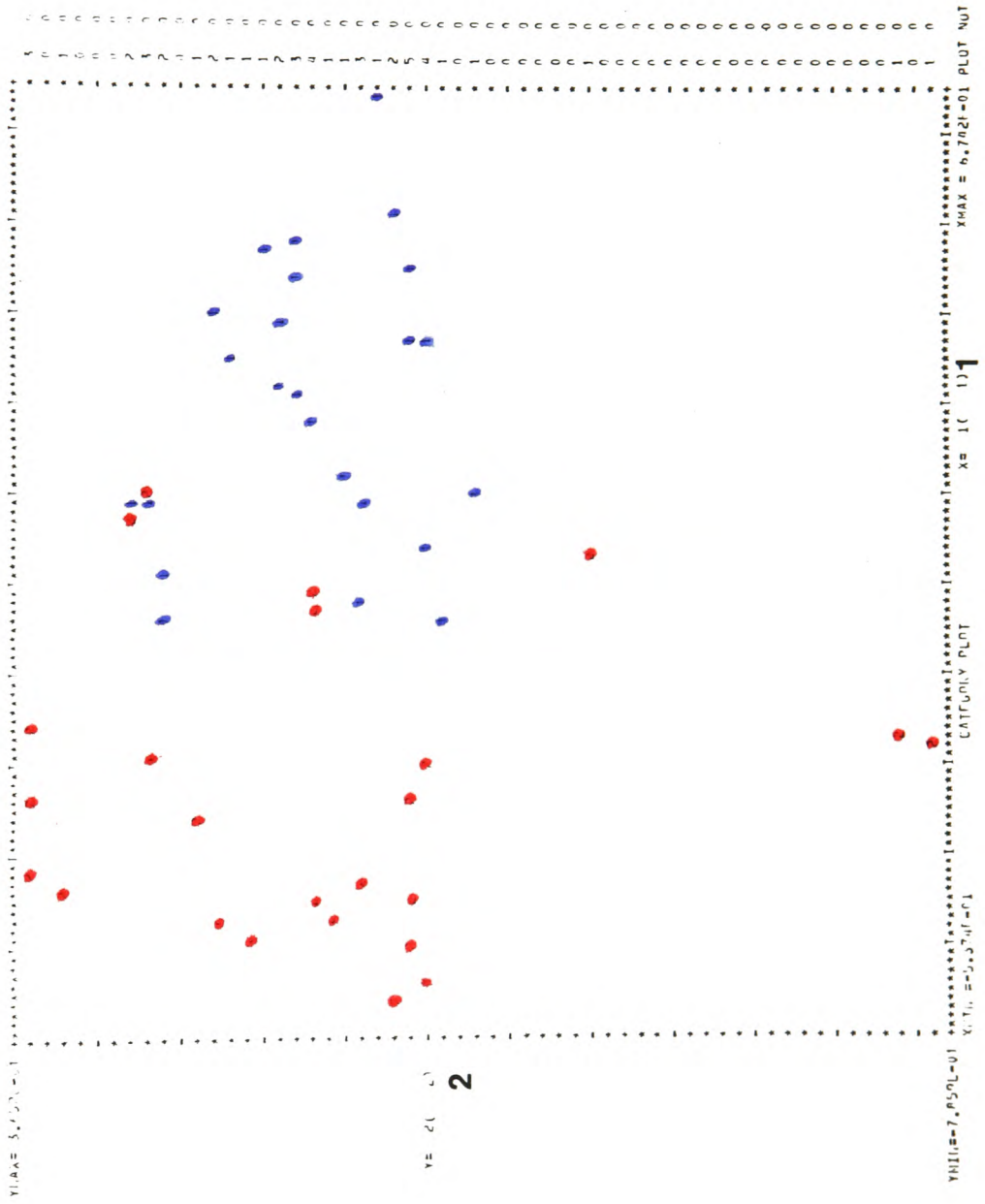


Figure 23. Principal component plot 1 v 2 for sites 3 and 4 only.



pointed out by Coomans and Masart (72). If two overlapping classes differ in size an object appearing in the overlap zone will probably be classified in the largest class, because in this zone there will be more neighbours of the largest class. Care must therefore be taken in interpreting results.

Table XVII shows a summary of results of the kNN analysis. It can be seen that no misclassification occurs for sites 1, 2 and 3 right out to the 10th nearest neighbour. It is only for category 4 that any misclassification occurs and then for only 1 sample up to the 4th nearest neighbour. In view of the fact that the number of samples collected for sites 3 and 4 were equal the misclassification is not due to a problem such as that pointed out by Coomans and Massart.

Additional data

Additional data was obtained from student projects and combined with the data from this study. This was then used in some pattern recognition methods to test the validity of the conclusions drawn above.

This additional data consisted of seven heavy metal concentrations on sediments from eight sampling sites. Six of these were from the Cynon, the remaining two from the Rhondda. Four of the Cynon sites were in fact at the same locations as the main study. These are shown in the

Table XVII **kNN Analysis summary.**

********** 1-NN ***

MISCLASSIFICATION MATRIX

COMPUTED CLASS / TRUE CLASS	1	2	3	4
1	50 100.0	0 0.0	0 0.0	0 0.0
2	0 0.0	48 100.0	0 0.0	0 0.0
3	0 0.0	0 0.0	24 100.0	0 0.0
4	0 0.0	0 0.0	1 4.5	21 95.5

PERCENT CORRECT, TOTAL..... 99.3
AVERAGE... 98.8

********** 3-NN ***

MISCLASSIFICATION MATRIX

COMPUTED CLASS / TRUE CLASS	1	2	3	4
1	50 100.0	0 0.0	0 0.0	0 0.0
2	0 0.0	48 100.0	0 0.0	0 0.0
3	0 0.0	0 0.0	24 100.0	0 0.0
4	0 0.0	0 0.0	1 4.5	21 95.5

PERCENT CORRECT, TOTAL..... 99.3
AVERAGE... 98.8

********** 4-NN ***

MISCLASSIFICATION MATRIX

COMPUTED CLASS / TRUE CLASS	1	2	3	4
1	50 100.0	0 0.0	0 0.0	0 0.0
2	0 0.0	48 100.0	0 0.0	0 0.0
3	0 0.0	0 0.0	24 100.0	0 0.0
4	0 0.0	0 0.0	1 4.5	21 95.5

PERCENT CORRECT, TOTAL..... 99.3
AVERAGE... 98.8

********** 5-NN ***

MISCLASSIFICATION MATRIX

COMPUTED CLASS / TRUE CLASS	1	2	3	4
1	50 100.0	0 0.0	0 0.0	0 0.0
2	0 0.0	48 100.0	0 0.0	0 0.0
3	0 0.0	0 0.0	24 100.0	0 0.0
4	0 0.0	0 0.0	3 13.0	19 86.4

PERCENT CORRECT, TOTAL..... 97.9
AVERAGE... 96.6

Table XVII **kNN Analysis summary continued.**

***

Table XVII **kNN Analysis summary continued.**

********** 10-NN ***

MISCLASSIFICATION MATRIX

COMPUTED CLASS / TRUE CLASS	1	2	3	4
1	50 100.0	0 0.0	0 0.0	0 0.0
2	0 0.0	48 100.0	0 0.0	0 0.0
3	0 0.0	0 0.0	24 100.0	0 0.0
4	0 0.0	0 0.0	3 13.6	19 86.4

PERCENT CORRECT, TOTAL..... 97.9
AVERAGE... 96.6

following Table XVIII.

TABLE XVIII. **Equivalent sites for main study and additional data.**

ARTHUR categories

Main study	Additional data
1	5
3	8
4	9
2	10

Figures 24 and 25 are principal component plots and feature-feature plots respectively. The similarity between the two plots is very obvious showing the importance of magnesium levels in differentiating between the sites. Another significant point to note is that despite overlap all categories are reasonably resolved even though some of them belong to the same geographical location. This demonstrates the effect of the time scale involved, The additional data having been collected about one year earlier than the main data. The striking feature is the difference between categories 8,9 and categories 3,4 their geographical eqivalents. This phenomenon may be accounted for by seasonal changes within the river system.

Figure 24. Principal component plot 1 v 2 including additional data.

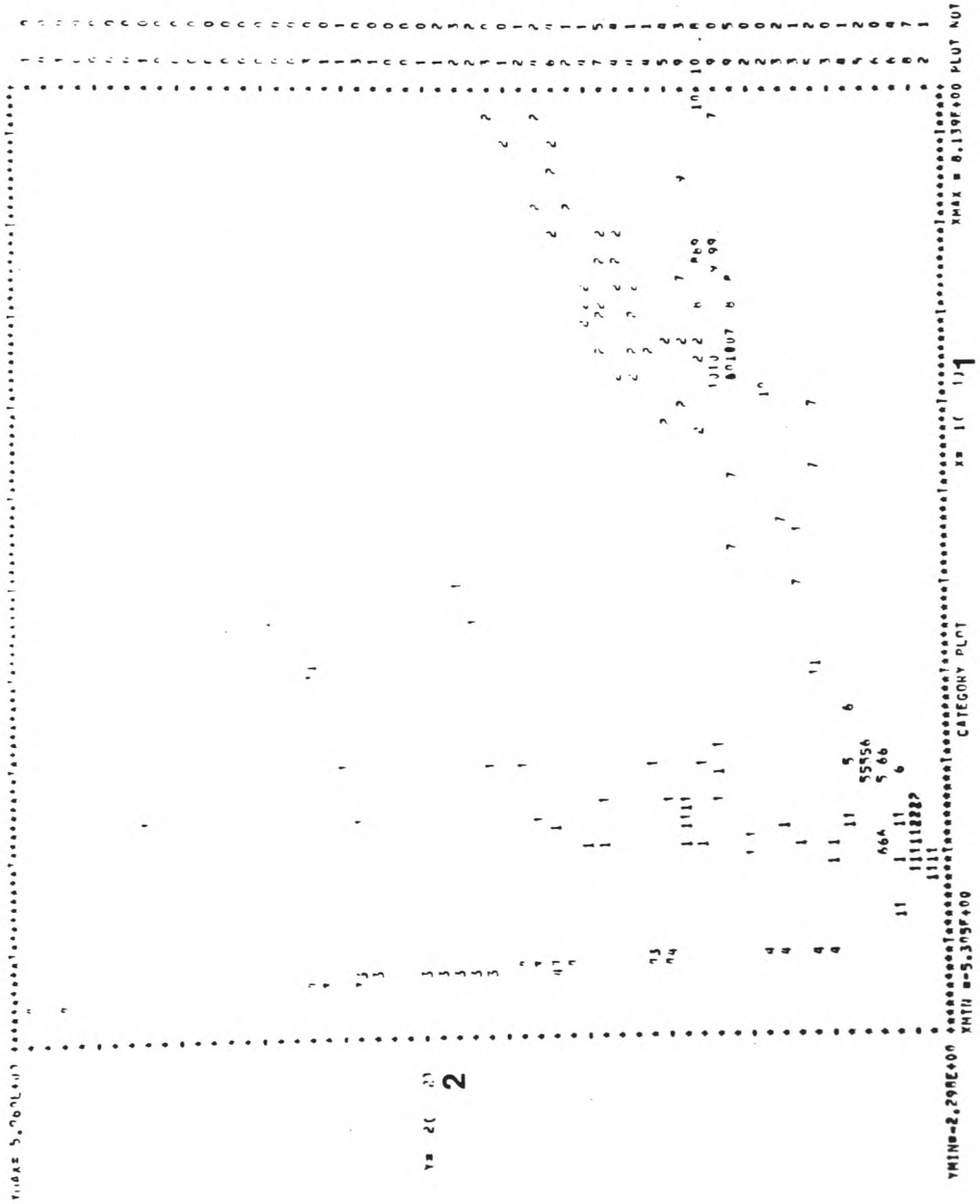
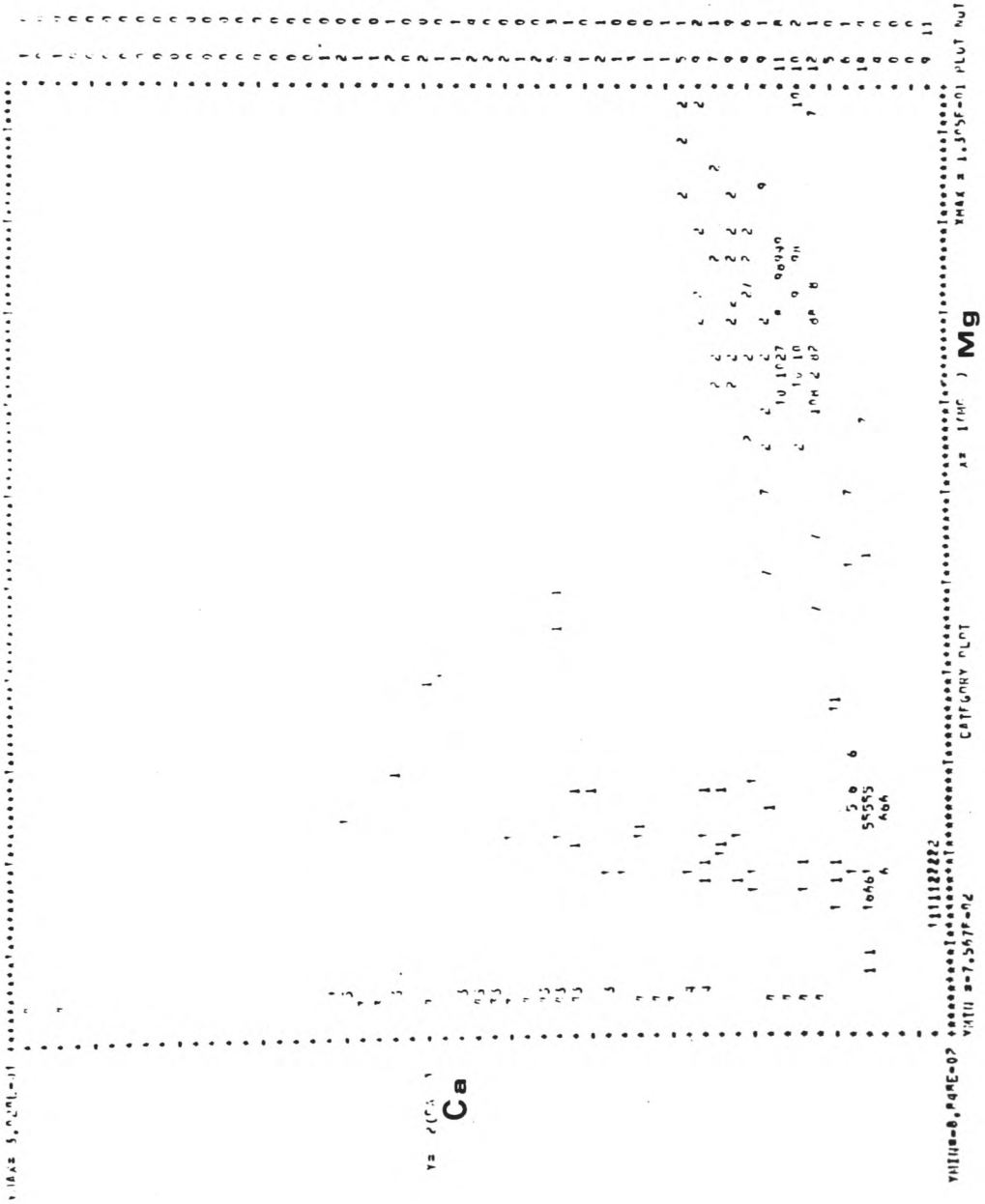


Figure 25. Feature-feature plot Mg v Ca including additional data.



A kNN analysis with the additional data placed in the test set correctly classified catagory 5 as category 1. Categories 6 and 7 were also classified as site 1. This is not entirely surprising as the sites are not too far apart. What was unexpected was the classification of category 10 (site 2 geographically) as site 1. This is a further demonstration of the dramatic changes that can take place in the sediments with time. Interestingly categories 11 and 12 (from a different river) were classified as site 4. However the significance of this is not clear as kNN **will** classify categories regardless of distance.

DISCUSSION

River sediment was sampled at four sites on the River Cynon. The samples were dried and divided into different size fractions. The smallest size fraction (<75um) was digested in a hydrochloric acid / hydrogen peroxide mixture. This method of digestion was chosen as a more rapid method than the more usual nitric / perchloric acid digestion after a small comparative study did not reveal any major differences between the methods. The concentrations of eleven heavy metals were determined for each sample in duplicate. The organic content of each sample was also estimated using the method of ignition loss.

In order to obtain an overview of trends within the river on moving downstream, the means of the metal concentrations were calculated for each site and these values were plotted. Rather surprisingly the highest values of some of the metals occurred at the uppermost site in the river. This site was on open moorland and free from any obvious source of pollution. It was concluded that the high values of Cd, Cr, Pb, Ni and Zn could only be as a result of the inherent geology of the sampling site. This hypothesis was supported by the high correlations observed between several metals at this site. Correlations between metals at the other sites were generally very low. The increase in Cu concentrations downstream suggest that there is significant input of this metal into the river. A possible source for this is a

wire and cable manufacturing industry in Aberdare which may also account for the increased aluminium levels at site 3.

The increase in organic content on moving downstream may well be due to coal based material. Typically coal from this area contains about 0.3% magnesium and the high level of magnesium at site 2 may therefore be due to this. Furthermore large volumes of minewater from both operational and closed pits are discharged into the Cynon (in the order of 2×10^6 gallons/day) which contain a significant concentration of magnesium. A heavy metal profile of coal (Appendix E) was compared to the target vector produced for site 2 by the factor analysis and found to be similar except for aluminium concentrations. These had been determined in the coal profile by neutron activation and would thus include aluminium combined in alumino-silicates. The digestion procedure used in this study would not have released aluminium bound in this way and so a second comparison was made without the aluminium concentrations. The two profiles were found to have a correlation coefficient $r = 0.88$. Similar values were determined for sites 3 and 4 suggesting that they too are closely related to coal.

The data was analysed using both Factor Analysis and Pattern Recognition techniques. The aim of the factor analysis was to determine the number of factors contributing to the heavy metal distribution at each site and if possible to obtain a profile of each factor. Pattern recognition methods were used to determine whether

sampling sites could be distinguished from each other.

The factor analysis identified one major factor which determined the heavy metal concentration at each site. The profile generated for site 1 was different from the others. Sites 3 and 4 were shown to be very similar (correlation coefficient r = 0.997) to each other and apart from magnesium very similar to site 2.In view of the geographical relationships of the sampling sites this is not an unexpected result.

At all of the sites except for site 2 the analysis indicated that second and possibly even third or fourth factors were also contributing to the heavy metal load in the sediments. Unfortunately, the analysis was unable to uniquely identify these factors and it is suggested that this may be due to the effect of analytical error being of the same order of magnitude as the secondary inputs.

It was therefore concluded, on the basis of the factor analysis that near its source the heavy metals in the Cynon are largely controlled by the local geology. Further downstream the heavy metal profiles are closely related to coal. This is not unexpected in view of the history of the Cynon valley. In the past it was dominated by mining activity and while much of that mining activity has now ceased it continues to have an impact through discharge of mine waters, runoff from spoil heaps and the presence of the Phurnacite works.

The pattern recognition techniques used were all capable of distinguishing between the sites with varying degrees of success. The usupervised learning methods of

hierarchical clustering resolved the data into 6 major clusters. Three of these were groups of samples belonging to site 1, one of these clusters containing 4 samples which appeared as outliers in virtually all of the other pattern recognition methods. One cluster contained all but two of the site 2 samples. One of the remaining clusters contained mainly site 3 samples with two site 4 samples while the other consisted of site 4 samples but included some site 3 samples.

This pattern of classification of the sites, identifying outliers and overlap of sites 3 and 4 was consistently repeated by the other unsupervised techniques.

Principal component plots correctly classified all the sites and identified the outliers at each site. Sites 3 and 4 were poorly resolved using the major principal components but were more successfully resolved using components from lower in the order. It is suggested that this is due to the variation at sites 1 and 2 being greater than for sites 3 and 4. This results in the major components being governed by the very different sites 1 and 2 and thus causing poor resolution of the similar sites 3 and 4. Plotting of the minor components does not produce as wide a separation of sites 1 and 2 but does achieve a better separation of sites 3 and 4. The selected feature-feature plots showed a similar sort of pattern. It was originally thought that this may be due to one particular element, magnesium; however, removal of this feature still produced a similar plot. This supports

the conclusion that the feature selected as producing the best classification is heavily weighted in favour of the dissimilar sites 1 and 2. Removal of sites 1 and 2 from the data matrix produced a good separation of sites 3 and 4. However the significance of magnesium has been demonstrated by the plotting methods used. It was selected as the metal best able to differentiate between the sites. After removal of sites 1 and 2 to enable greater resolution between sites 3 and 4, it was still magnesium that was selected as the feature with the greatest ability to classify the sites.

Non-linear mapping produced quite distinct groups with little overlap and outliers were once again identified. In terms of achieving a good resolution of the four sites this particular technique was found to be the most successful but apart from the identification of the outliers little other information can be obtained.

The supervised learning technique of k-nearest-neighbour classified sites 1,2 and 3 with 100% success out to the tenth nearest neighbour with only 1 site 4 sample missclassified as site 3 up to the fourth nearest neighbour. The use of additional data showed that classification is still possible on data taken from the same site at a different time. This demonstrates the ability of pattern recognition to recognise changes with time and highlights the continuous redistribution of the heavy metal profile in the sediment.

Pattern recognition techniques have been shown to be useful in classifying environmental data from river

systems. The methods are sensitive enough to identify outliers or unique samples within the data and to resolve very similar sites. Magnesium has been shown to be an important element in discriminating between sites within the river.

CONCLUSIONS

The survey of four sites in the River Cynon showed the river to contain levels of heavy metals significantly higher than some comparative, unmineralised rivers such as the Fowey in Cornwall. However, the Carnon, another Cornish river, heavily mineralised through mining activity, had somewhat higher levels than the Cynon.

The Cynon may therefore be considered to contain sediments typical of an industrialised mining valley with heavy metal concentrations higher than rivers not subjected to such influences but at the same time there is little evidence of very heavy pollution.

The sampling site near the source of the river had surprisingly high levels of heavy metals in the sediment which appear to be due to the inherent geology of the site. The sampling sites further downstream had considerably lower values of Cd, Cr, Pb, Mn, Ni and Zn. There was a slight rise in mean values at the lowermost sampling site. Copper concentrations increased downstream suggesting that there is input of this metal into the river which is possibly as a result of a wire and cable industry in the valley. This might also account for the increase in aluminium concentration. Calcium levels increase between the first and second sampling sites and decrease again downstream. The river passes over a bed of carboniferous limestone between these two sites which probably accounts for this.

The organic content of the river as determined by loss

on ignition increased downstream with an average of almost 32% organic content in the sediment of the lowest sampling site in the river. This is suspected of being very largely due to the presence of coal and allied minerals in the sediment. This postulate is supported by the visual inspection of the sediments collected at each site, although no definitive analysis of the mineral content of the samples was attempted. The major evidence in support of this is the comparison of the heavy metal profile of coal with the target vectors predicted for site 2.

The factor analysis of the data obtained showed that one major factor was controlling the heavy metal distribution at each site. At the lowest site in the river, all the indications of the size of the factor space point to one factor only whereas at the other sites second or even third and fourth factors may be making a contribution.

The profiles of the major factor at each site were generated using target factor analysis. The profiles of sites 2,3 and 4 were shown to be similar while that for site 1 was appreciably different. It proved impossible to identify any of the minor factors as several were generated and no satisfactory means of distinguishing them could be found.

Pattern recognition techniques were found to classify heavy metal concentrations for the samples into groups attributable to each site. Separation of site 1, site 2, sites 3 and 4 were good but most methods could not resolve completely site 3 and site 4. However satisfactory

resolution of these very similar sites can be achieved by utilising plots of lower order principal components. This is a demonstration of the ability of these methods to identify adequately, two very similar groups of samples as different classes.

To my knowledge there has been no other attempt at using factor analysis and pattern recognition techniques on data matrices of heavy metals in river sediments. The usefulness of pattern recognition in site classification has been demonstrated. Factor analysis can be used to determine the number of factors controlling heavy metal concentrations at each site and to obtain profiles of those factors.

A possible extension of the work in this area would be to discover the minimum number of sites necessary to classify the whole river. Further refinements would include determining the minimum number and identity of those elements needed to classify satisfactorily each site. This would provide an excellent background from which to monitor heavy metal distribution over a longer period and a variety of seasonal changes as well as comparisons with other river systems.

BIBLIOGRAPHY

1. C Herschell in 'Frontinus and the Water Supply of the City of London'. Longmans London p67 (1913)
2. Encyclopaedia of Chemical Science. Van Nostrand p533 Princeton (1964)
3. D N Lapedes Dictionary of Scientific and Technical Terms. McGraw Hill New York p674 (1974)
4. D C Burrel in 'Atomic spectrometric analysis of heavy metal pollutants in water' Ann Arbor Science Michigan p19 (1974)
5. E Nieberer, D H S Richardson Environ. Poll.Series B 1 p3 (1980)
6. S R Aston, I Thornton. Sci. Total Environ. 7 247(1977)
7. B E Brown, Hydrobiologia (Neth.) 52 221 (1977)
8. J D Parsons, Water Air Soil Pollut. 7 333 (1977)
9. P L Rose, Public Health Engineer. 13 167 (1985)
10. R M Harrison, S J Wilson. Sci. Total Environ. 43 63 (1985)
11. J Pigg. Proc. Okla. Acad. Sci. 55 19 (1975)
12. B McDuffie, I El-Barbery, G J Hollod, R D Tiberio. Trace Substances Environ. Health 10 85 (1976)
13. R Wollast. Water Sci. Tech. 14 107 (1982)
14. R Petersen, Environ. Sci. Tech. 16 443 (1982)
15. H E Allen, R H Hall, T D Brisbin. Environ. Sci. Tech. 14 441 (1980)
16. J Buffle. Trends in Analytical Chemistry. 1 90 (1981)
17. T M Florence. Talanta 29 345 (1982)

18. C Chakoumakos, R C Russo, R V Thurston. Environ. Sci. Tech. 13 213 (1979)

19. J A Davies, J O Leckie. Environ. Sci. Tech. 12 12 (1978)

20. M Astruc, J Lecomte, P Mericam. Environ. Tech. Lett. 2 1 (1981)

21. D P H Laxen, R M Harrison. Water Res. 15 1053 (1981)

22. C Mouvet, A C M Bourg. Water Res. 17 641 (1983)

23. S W Effler, M C Rand, T A Tamayo. Water Air Soil Pollut. 12 2 (1979)

24. D P H Laxen, E R Sholkovitz. Environ. Tech. Lett. 2 12 561 (1981)

25. H Bloom, G M Ayling. Environ. Geol. 2 3 (1977)

26. E A Jenne. Adv. Chem. Series 73 (1968)

27. J W Moore. Bull. Environ. Contam. Toxicol. 24 6 828 (1980)

28. R C Sidle, L T Kardos. Water Air Soil Pollut. 8 453 (1977)

29. J Gardiner. Water Res. 10 507 (1976)

30. J Slavek, W E Pickering. Water Air Soil Pollut. 16 209 (1981)

31. P S Rendell, G E Batley, A J Cameron. Environ. Sci. Tech. 14 3 314 (1980)

32. J Vuceta, J J Morgan. Environ. Sci. Tech. 12 12 1302 (1978)

33. M M Benjamin, J O Leckie. Environ. Sci. Tech. 15 9 1050 (1981)

34. M M Benjamin, J O Leckie. Environ. Sci. Tech.16 162 (1982)

35. H M N H Irving. Dithizone. Chem. Soc. London p50 (1977)

36. K Burger. Organic Reagents in Metal Analysis. Pergamon Press Oxford. p122 (1973)

37. I L Marr, M S Cresser. Environmental Chemical Analysis International Textbook Co. Glasgow. (1983)

38. A Walsh. Spectrochim. Acta 7 108 (1955)

39. J Wang. Environ. Sci. Tech. 16 104A (1982)

40. H W Nurnberg, P Valenta, L Mort, B Raspor, L Sipos. Z. Anal. Chem. 282 357 (1976)

41. M Tsurumi. Analyt. Chim. Acta 138 177 (1982)

42. E R Malinowski, D G Howery. Factor Analysis in Chemistry. Wiley Interscience NewYork (1980)

43. D L Massart, L Kaufman. The Interpretation of Analytical Chemical Data by the use of Cluster Analysis. J Wiley & Sons New York 1983)

44. E J Knudson, D L Duewer, G D Christian, T V Larson. in Chemometrics: Theory and Applications.Kowalski (ed) American Chemical Soc. Washington D C. p80 (1977)

45. W J Dunn, D L Stalling, J W Hogan, J D Petty, E Johansson, S Wold. Anal. Chem. 5 1308 (1984)

46. D F Gatz. J. Appl. Met. 17 600 (1975)

47. P K Hopke. J. Environ. Sci. Health A11(6) 367 (1976)

48. I H Blifford, G O Meeker. Atmospheric Environ. 1 147 (1967)

49. E Jellum, I Bjornson, R Nesbakken, E Johansson, S Wold J. Chromatogr. 217 231 (1981)

50. W J Dunn, S Wold, V Edlund, J Gasteiger, S Hellberg. Kemia-Kemi 9 401 (1982)

51. P K Hopke, E S Gladney, G E Gordon, W H Zoller, A G Jones. Atmosheric Environ. 10 1015 (1976)

52. P K Hopke. Proc. Air Pollution Control Assoc. 4 1 (1981)

53. P D Gaarenstroom, S P Perone, J L Mayers. Environ. Sci. Tech. 11 600 (1977)

54. C W Lewis, E S Macias. Atmospheric Environ. 14 185 (1980)

55. B Bobee, M Lachance, O Cazaillet. Eau de Quebec. 16 56 (1983)

56. F Tada, H Nishida, M Miyan, S Suzuki. Environ. Geol. 4 217 (1983)

57. H Nishida, F Tada, S Suzuki. Report of Research Nippon Institute of Technology. 10 87 (1980)

58. T J Hoban. Ph D. Thesis CNAA Polytechnic of Wales. (1985)

59. T N George, in British Regional Geology - South Wales. H.M.S.O. (1970)

60. The Wolfson Geochemical Atlas of England and Wales. Clarendon Press Oxford (1978)

61. B G Oliver. Environ. Sci. Tech. 7 135 (1973)

62. J O Nriagu, A W Kemp, H K T Wong, N Harper. Geochim. Cosmochim. Acta. 43 247 (1979)

63. T D Rees,J Hilton. Lab. Practice. 27 291 (1978)

64. W E Dean. J. Sediment Petrol. 44 242 (1974)

65. R J Rummel Applied Factor Analysis. Northwestern University Press Evanston Ill. (1970)

66. W B Lyons, P A Mayewski,F H Chormann. Chemical Geol. 48 265 (1985)

67. S R Aston, I Thornton, J S Webb, J B Purves, B L Milford. Water Air Soil Pollut. 3 321 (1974)

68. ARTHUR 81 Infometrix Inc. Seattle (1981)

69. B R Kowalski, C F Bender. J Am. Chem. Soc. 94 5632 (1972)

70. J W Sammon. IEEE Trans Computers, C18 401 (1969)

71. B R Kowalski, C F Bender. J. Am. Chem. Soc. 95 686 (1973)

72. D Coomans, D L Massart. Analytica Chimica Acta 138 15 (1982).

APPENDIX A

Comparison of digestion methods.

Relative extraction efficiency A.

Relative extraction efficiency B.

COMPARISON OF DIGESTION METHODS

	Cu	Cd	Cr	Ni	Zn	Pb
SAMPLE A						
Hydrochloric acid/ hydrogen peroxide	9.58	0.49	53.4	16.5	57.1	49.2
Perchloric acid/ nitric acid	8.20	0.73	46.9	18.7	62.0	51.7
SAMPLE B						
Hydrochloric acid/ hydrogen peroxide	57.7	0.83	18.9	37.4	152.7	100.5
Perchloric acid/ nitric acid	58.3	0.91	19.7	41.8	145.3	109.1

Metal Conc. v Mass Sample for HCl/H2O2 Digestion.
250
200
150
100
50
0
Metal Conc.
Zn
Pb
Cu
Ni
Cr
Cd
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Mass Sample (g)

Metal Conc. v Sample Mass for HCl/H2O2 Digestion
Zn
Cr
Pb
Ni
Cu
100
50
0
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
mass sample (g)

APPENDIX B

Atomic absorption spectroscopy instrument parameters.

metal	gas	lamp current	wavelength nm	slit width nm
Al	N-A	10	309.3	0.5
Cr	A-A	7	240.7	0.2
Cd	A-A	4	228.8	0.5
Cu	A-A	4	324.8	0.5
Fe	A-A	5	386.0	0.2
Mg	A-A	4	202.5	0.5
Mn	A-A	5	403.1	0.2
Pb	A-A	5	217.0	1.0
Zn	A-A	5	213.9	1.0
Ni	A-A	4	232.0	0.2

N-A : Nitrous oxide - acetylene

A-A : Air - acetylene

APPENDIX C

Results of metal determinations.

SITE	SAMPLE	Al	Ca	Cd	Cr	Cu	Fe	Pb	Mg	Mn	Ni	Zn
1	1	5710	6520	14.3	25.0	27.7	25000	158	429	5800	49.1	1210
1	2	5580	6350	15.4	27.9	30.8	25000	166	365	6150	53.8	1250
1	3	6580	7370	72.4	31.6	38.2	44800	222	395	21100	189.0	3030
1	4	8000	8130	42.5	32.5	33.7	48800	244	550	23800	119.0	3290
1	5	4320	2160	1.1	18.2	11.4	4770	27	409	227	12.5	76
1	6	4000	2200	1.0	18.0	13.0	5000	27	440	200	13.0	76
1	7	7220	7360	54.2	27.8	44.4	41700	293	556	29200	143.0	3920
1	8	8130	8440	56.3	28.1	53.1	43800	302	688	29700	145.0	4030
1	9	5470	4530	19.5	13.3	27.3	25800	189	375	9140	60.2	1740
1	10	5970	4860	20.1	13.9	32.6	25700	193	444	9030	60.4	1740
1	11	5930	7590	31.4	22.2	28.7	24100	194	685	15600	82.4	2350
1	12	5450	7500	31.8	21.6	27.3	23900	191	682	15300	79.5	2390

SITE	SAMPLE	Al	Ca	Cd	Cr	Cu	Fe	Pb	Mg	Mn	Ni	Zn
1	13	7690	7690	39.4	24.0	31.7	37500	279	500	20500	99.0	2880
1	14	7500	7790	40.4	26.9	31.7	36500	295	481	20300	99.0	2850
1	15	12200	2620	0.6	40.9	7.9	31100	47	1460	366	15.2	67.1
1	16	12800	2700	0.5	39.3	8.2	30600	48.5	1430	357	15.8	61.2
1	17	5900	7110	15.4	14.7	37.2	27600	163	551	6990	106.0	1060
1	18	6170	8330	15.8	16.7	33.3	27500	158	550	7920	42.5	1150
1	19	7750	6190	30.0	17.5	26.9	41300	204	638	15600	82.5	2170
1	20	7790	6450	32.0	20.3	30.2	41300	221	721	17400	87.8	2270
1	21	3250	2590	1.3	9.7	11.0	2500	23.2	298	175	7.5	140
1	22	3670	2500	1.1	8.9	10.6	2720	22.8	311	222	7.8	144
1	23	1790	2130	3.0	4.5	17.5	8580	94	119	1160	9.0	146
1	24	2400	2400	3.5	4.0	20.0	10500	106	160	1300	11.0	180

SITE	SAMPLE	Al	Ca	Cd	Cr	Cu	Fe	Pb	Mg	Mn	Ni	Zn
1	25	9210	10700	42.1	13.1	44.7	40800	266	579	2700	117.0	3200
1	26	8330	10700	41.7	13.9	34.7	40300	263	556	26400	118.0	3130
1	27	8520	8240	31.5	10.2	38.0	35200	222	537	16400	89.8	2510
1	28	7380	8100	29.8	11.9	39.3	33300	215	452	16400	81.0	2500
1	29	6830	3420	1.7	18.3	20.0	20800	29.2	983	667	22.5	175
1	30	7330	3420	1.7	20.0	20.0	20800	27.5	950	667	22.5	175
1	31	4830	8080	17.5	11.7	20.8	23300	159	400	11200	53.3	1400
1	32	4780	9020	18.5	8.7	21.7	22800	164	413	11700	54.3	1500
1	33	6350	11400	31.7	20.2	24.0	34600	221	423	16800	87.5	2420
1	34	6610	11700	30.6	19.4	25.8	32300	218	419	15300	83.9	2310
1	35	6670	13500	24.3	13.9	25.7	28500	198	528	12400	66.0	1820
1	36	5230	13100	25.0	13.6	25.0	29500	189	500	12500	64.8	1860

SITE	SAMPLE	Al	Ca	Cd	Cr	Cu	Fe	Pb	Mg	Mn	Ni	Zn
1	37	13200	13800	44.9	21.3	142.0	51500	322	1320	25100	122	3160
1	38	12500	13500	45.8	21.9	150.0	53100	325	1210	25300	124	3440
1	39	4550	3640	8.0	5.7	22.7	19300	140	386	2840	31.8	943
1	40	3570	3810	8.3	4.8	21.4	19000	136	310	2740	28.6	881
1	41	12700	18200	95.5	29.5	97.7	88600	613	1050	48900	295	6700
1	42	12700	17700	102.0	47.9	150.0	89600	623	1080	47700	429	6560
1	43	17500	15500	52.0	64.0	118.0	129000	216	540	28000	410	6000
1	44	15000	21500	60.0	40.0	85.0	163000	300	600	27500	390	8000
1	45	8000	12300	32.5	25.0	60.0	60800	150	700	15000	108	3500
1	46	9000	13300	37.5	25.0	62.5	68800	156	700	16600	108	3500
1	47	10500	19500	32.0	24.0	30.0	50200	92	760	20000	89	2700
1	48	9400	19500	34.0	19.0	32.0	49500	93	750	21000	101	2700

SITE	SAMPLE	Al	Ca	Cd	Cr	Cu	Fe	Pb	Mg	Mn	Ni	Zn
1	49	4600	3100	6.0	6.0	17.0	13600	49	440	760	23	520
1	50	4800	3300	6.0	6.0	17.0	14200	49	460	790	23	570
2	1	7590	4440	1.9	10.2	65.7	50200	83	2040	593	37	143
2	2	7120	4420	1.9	9.6	64.4	50800	82	2120	596	37	147
2	3	4800	5000	2.0	8.0	70	47000	99	1800	680	34	129
2	4	4800	5000	2.0	8.0	74	45000	102	1800	670	35	129
2	5	6150	7400	2.9	13.4	119	56700	142	2600	904	51	184
2	6	5800	7600	3.0	13.0	124	62000	147	2700	960	54	183
2	7	5600	7000	2.0	14.0	89	51500	118	2500	730	44	159
2	8	5600	6900	2.0	12.0	98	50500	118	2400	720	45	155
2	9	5000	6630	1.9	7.7	99	27400	107	1830	606	33	128
2	10	5000	6630	1.9	10.6	111	27900	116	1830	606	34	132

SITE	SAMPLE	Al	Ca	Cd	Cr	Cu	Fe	Pb	Mg	Mn	Ni	Zn
2	11	4400	6900	2	10	93	35000	120	2200	750	36	150
2	12	5400	7000	2	10	90	38500	115	2400	770	38	152
2	13	5000	6300	2	9	83	34000	99	2100	700	36	147
2	14	5000	6400	2	9	83	35500	97	2100	720	37	147
2	15	5200	7500	2	16	78	45000	103	2400	690	42	200
2	16	5200	7300	2	16	73	45000	100	2500	690	42	200
2	17	5400	6700	2	14	80	51500	103	2400	670	45	151
2	18	5800	6800	2	10	75	52000	99	2400	670	46	152
2	19	5200	6300	2	11	89	34500	98	2100	750	38	145
2	20	4800	6100	2	10	88	34000	100	2100	740	36	148
2	21	5600	5300	2	11	67	37000	92	2100	470	38	182
2	22	5800	5300	2	11	66	37000	90	2100	470	39	175

SITE	SAMPLE	Al	Ca	Cd	Cr	Cu	Fe	Pb	Mg	Mn	Ni	Zn
2	23	5200	6800	2	11	73	38000	92	2500	690	55	169
2	24	5000	6600	2	9	70	37500	90	2300	690	37	165
2	25	5200	7200	3	13	87	32500	112	2500	350	40	215
2	26	5200	6900	3	13	81	33000	110	2500	340	40	214
2	27	5600	8600	2	15	84	46000	108	2900	800	41	165
2	28	5600	8700	2	15	92	45500	107	2800	800	42	163
2	29	4600	7100	2	15	316	36000	268	2000	730	35	169
2	30	4200	6800	2	13	282	35500	244	2100	700	34	164
2	31	3400	6200	1	5	52	50500	81	2200	700	38	136
2	32	4000	6500	1	9	54	51500	79	2300	680	38	133
2	33	4500	7000	2.5	11	84	58100	120	2250	638	44	150
2	34	4500	7130	2.5	11	78	56300	115	2250	650	44	148

SITE	SAMPLE	Al	Ca	Cd	Cr	Cu	Fe	Pb	Mg	Mn	Ni	Zn
2	35	4400	9100	2	14	110	74000	134	2900	850	47	160
2	36	4400	9200	2	15	105	72000	128	2800	840	48	156
2	37	4400	8200	2	15	103	48000	123	2300	870	44	170
2	38	4800	8200	2	12	99	46000	119	2200	870	44	175
2	39	4400	7100	2	13	97	64500	121	2100	770	46	152
2	40	4000	7100	2	11	89	64000	109	2200	760	45	152
2	41	4200	5800	2	7	66	26500	81	1800	660	31	133
2	42	4800	6200	2	9	70	28000	88	1900	680	35	142
2	43	5400	8700	2	17	73	56500	102	2600	770	46	164
2	44	5200	8500	2	15	79	56000	104	2500	740	47	166
2	45	4200	7500	2	7	112	49000	128	2000	760	44	155
2	46	4000	7700	2	11	107	48500	128	2100	750	44	149

SITE	SAMPLE	Al	Ca	Cd	Cr	Cu	Fe	Pb	Mg	Mn	Ni	Zn
2	47	4500	7250	2.5	9	88	63800	129	2250	700	50	160
2	48	4500	7380	2.5	9	91	65000	131	2250	713	54	161
3	1	8000	9600	0	6	104	50000	68	28	1080	25	177
3	2	9600	9400	0	8	112	50000	67	29	1110	25	206
3	3	10400	17300	0	15	72	49500	74	36	1060	31	251
3	4	9700	17400	0	13	71	49000	79	36	1060	33	249
3	5	8200	21400	1	8	62	43000	134	33	820	23	194
3	6	8400	19800	2	8	71	42000	154	32	810	48	219
3	7	10400	12200	2	14	82	62500	108	36	1420	77	279
3	8	10400	12200	1	13	79	60000	98	36	1400	44	261
3	9	10400	16500	1	15	79	49500	77	36	1060	45	251
3	10	11100	17000	1	15	83	50000	58	35	1130	45	261

SITE	SAMPLE	Al	Ca	Cd	Cr	Cu	Fe	Pb	Mg	Mn	Ni	Zn
3	11	9700	13400	0	8	47	42000	66	32	910	30	203
3	12	10800	13400	0	11	47	42000	69	33	910	30	202
3	13	10100	14300	1	12	64	44000	147	33	1080	36	223
3	14	10400	13600	1	14	68	45000	131	33	1110	38	219
3	15	8000	16200	1	12	63	39500	114	30	1110	35	251
3	16	8500	15600	1	12	59	40000	107	31	1110	34	267
3	17	9700	13800	1	13	82	43000	90	35	1110	35	244
3	18	10400	14000	1	15	86	44000	82	35	1120	39	249
3	19	9800	16800	1	12	159	41000	163	34	750	40	293
3	20	8800	16600	1	12	149	39500	165	34	750	38	292
3	21	9200	16400	2	15	78	42500	109	33	1180	42	278
3	22	10000	16900	2	15	80	41500	104	33	1140	41	269

SITE	SAMPLE	Al	Ca	Cd	Cr	Cu	Fe	Pb	Mg	Mn	Ni	Zn
3	23	8400	20500	2	12	57	38000	135	31	770	36	207
3	24	7400	20600	2	10	55	37000	131	31	750	33	201
4	1	9600	8600	1	10	50	52000	58	32	750	29	223
4	2	9300	8900	1	12	47	50500	60	32	720	35	221
4	3	9600	6000	0	8	43	37000	44	25	680	19	192
4	4	8600	5700	0	8	37	36500	47	23	670	17	181
4	5	8100	4100	0	5	37	36500	49	18	980	16	180
4	6	9000	3900	0	5	39	35500	49	19	940	16	171
4	7	6100	13300	2	6	36	24000	59	24	1360	16	135
4	8	9200	22200	2	6	61	33000	110	34	2300	33	238
4	9	6300	31800	3	5	67	24500	82	30	3300	32	208
4	10	5900	32800	3	5	71	24000	84	30	3400	31	208

SITE	SAMPLE	Al	Ca	Cd	Cr	Cu	Fe	Pb	Mg	Mn	Ni	Zn
4	11	7500	13800	3	5	48	24000	72	21	770	21	205
4	12	7200	14200	2	5	51	22000	73	21	780	21	211
4	13	6900	13800	2	5	47	33500	184	28	820	23	202
4	14	6600	14800	2	5	44	34500	154	28	840	22	191
4	15	7100	5200	2	5	47	39000	60	23	400	21	222
4	16	9400	5200	2	5	44	39500	61	24	400	20	226
4	17	8200	4500	2	5	41	29000	128	17	550	18	184
4	18	7100	4700	2	5	36	31000	144	18	540	17	183
4	19	6700	18200	2	8	38	23000	82	20	1100	17	162
4	20	5600	17000	2	7	39	21500	72	20	1100	17	172
4	21	8500	9400	2	10	76	44000	175	24	780	30	
4	22	8500	8600	2	11	78	44000	173	24	770	31	

SITE	SAMPLE	Al	Ca	Cd	Cr	Cu	Fe	Pb	Mg	Mn	Ni	Zn
4	23	6900	10700	2	7	45	24500	54	19	830	23	203
4	24	5900	9900	2	7	51	24000	51	19	810	25	197

APPENDIX D

Percentage organic content as determined by ignition loss.

SITE 1

SAMPLE No	% ORGANIC
1.01	12.18
1.01	11.57
1.02	11.56
1.02	10.75
1.03	4.87
1.03	4.83
1.04	11.03
1.04	15.57
1.05	11.52
1.05	11.39
1.06	9.55
1.06	10.20
1.07	9.01
1.07	8.76
1.08	4.29
1.08	4.26
1.09	10.93
1.09	10.80
1.10	11.27
1.10	10.98
1.11	7.15
1.11	7.17
1.12	9.81
1.12	9.51
1.13	11.42
1.13	11.11
1.14	11.60
1.14	11.13
1.15	3.76
1.15	3.75
1.16	9.01
1.16	8.69
1.17	7.70
1.17	7.46
1.18	9.99
1.18	8.69
1.19	10.27
1.19	9.31
1.20	8.49
1.20	8.04
1.21	5.96
1.21	5.94
1.22	***
1.23	9.48
1.23	***
1.24	9.87
1.24	***
1.25	6.80
1.25	6.61

SITE 3

SAMPLE No	% ORGANIC
3.01	13.56
3.01	12.24
3.02	12.98
3.02	12.45
3.03	13.18
3.03	11.74
3.04	11.70
3.04	11.41
3.05	11.77
3.05	11.55
3.06	11.13
3.06	11.12
3.07	12.57
3.07	12.67
3.08	16.65
3.08	16.95
3.09	13.61
3.09	13.41
3.10	12.90
3.10	12.62
3.11	12.51
3.11	14.15
3.12	12.57
3.12	12.56

*** - insufficient sample

SITE 4

SAMPLE No	% ORGANIC
4.01	13.71
4.01	13.74
4.02	18.01
4.02	17.57
4.03	21.07
4.03	21.09
4.04	19.43
4.04	19.23
4.05	25.76
4.05	24.58
4.06	17.02
4.06	17.85
4.07	14.13
4.07	14.10
4.08	15.74
4.08	15.55
4.09	10.43
4.09	10.18
4.10	29.96
4.10	29.35
4.11	17.20
4.11	17.14
4.12	16.23
4.12	18.92

SITE 2

SAMPLE No	% ORGANIC
2.01	26.83
2.01	26.93
2.02	35.99
2.02	36.26
2.03	***
2.04	21.74
2.04	22.34
2.05	24.63
2.05	24.75
2.06	47.12
2.06	47.30
2.07	32.12
2.07	32.02
2.08	38.71
2.08	38.74
2.09	15.64
2.09	15.79
2.10	20.81
2.10	21.30
2.11	40.27
2.11	40.16
2.12	15.92
2.12	16.36
2.13	32.29
2.13	32.06
2.14	34.03
2.14	33.62
2.15	34.82
2.15	34.32
2.16	42.59
2.16	42.60
2.17	6.55
2.17	6.31
2.18	34.66
2.18	36.92
2.19	32.45
2.19	31.79
2.20	34.72
2.20	35.02
2.21	39.64
2.21	41.98
2.22	43.14
2.22	43.18
2.23	28.62
2.23	28.25
2.24	44.42
2.24	46.49
2.25	31.07
2.25	33.96

*** - insufficient sample

APPENDIX E

Typical heavy metal profile of coal.

Heavy metal concentrations expressed as $\mu g\ g^{-1}$ unless stated otherwise.

Al	3.0%
Ca	0.2%
Cd	<0.3
Cr	40
Cu	35
Fe	0.5%
Pb	10
Mg	0.26%
Mn	50
Ni	35
Zn	33